الحاسب الآلى

التحليل الاحصائى

باستخدام برنامج SPSS

مقدمة:

يعد برنامج Statistical Package for Social Science (SPSS) أقدم البرامج الإحصائية وأكثرها استخداماً من قبل شريحة واسعة من الطلبة والباحثين في مختلف التخصصات الإحصائية والطبية والهندسية والزراعية و الاجتماعية والتربوية والنفسية. ونظراً لقلة عدد من يجيد استخدام برنامج الـ SPSS بصورة وافية إضافة إلى افتقار المكتبة العربية إلى كتب تعليمية حول هذا البرنامج فقد كان هدفنا ومن خلال هذا الجزء أن نضع مصدراً تفصيلياً بين أيدي الباحثين في مختلف المجالات وطلاب مرحلة البكالوريوس في أقسام الإحصاء والتخصصات المختلفة الاخرى.

يتضمن هذا الجزء تعريفاً بالجوانب الأساسية لحزمة SPSS ويهدف إلي إكساب العديد من المهارات اللازمة لتحقيق الاستفادة القصوى من إمكانيات البرنامج المتاحة آخذين في الاعتبار أن هناك عدد كبيراً من الدارسين ليس لديهم خلفية إحصائية وافية تمكنهم من التعامل مع البرنامج بصورة صحيحة.

يتضمن هذا الجزء نواحي ادارة الملفات و استخدام البرنامج كقاعدة بيانات فيما يتعلق بدمج الملفات وترتيبها واختيار الحالات وتبادل البيانات مع البرامج الأخرى كذلك الرسوم البيانية والجداول التكرارية والمقاييس الوصفية. كما يتضمن عرضاً موجزاً للجانب النظري للأسلوب الإحصائي المستخدم بالإضافة إلى التفسير الإحصائي لمخرجات البرنامج لبعض التطبيقات الإحصائية المهمة مثل اختبارات الفرضيات، الاختبارات المعلمية و اللامعلمية، تحليل التباين ، الانحدار...الخ. وقد تم التعامل مع تطبيقات البرنامج من خلال أمثلة مبسطة تتيح للقارئ الانتقال إلى خطوات متقدمة بسهولة ومعظم هذة الأمثلة مأخوذ من مصادر عربية وأجنبية معتمدة.

د. مهدى محمد القصاص

مقدمة لبرنامج SPSS

التطور التاريخي:

من المعلوم أنه لحساب قيمة أي مقياس إحصائي فإننا بحاجه إلى إجراء العديد من الخطوات الرياضية. على سبيل المثال لحساب قيمة الانحراف المعياري (أحد مقاييس التشتت) فإننا نقوم باجراء العمليات الحسابية التالية:

١. ايجاد عدد المشاهدات،
٢. حساب مجموع المشاهدات،
٣. قسمة قيمة ناتج العملية الحسابية رقم (٢) على عدد البيانات والذي يعطي قيمة الوسط الحسابي،
٤. لكل مشاهدة يتم إيجاد الإنحراف عن الوسط الحسابي المحسوب في العملية الحسابية رقم (٣)،
٥. حساب مربع قيمة ناتج العملية الحسابية رقم (٤) لكل حالة،
٦. ايجاد مجموع مربعات قيمة ناتج العملية الحسابية رقم (٥)،
٧. قسمة قيمة ناتج العملية الحسابية رقم (٦) على عدد المشاهدات إذا كنا نتعامل مع بيانات المجتمع أو القسمة على عدد المشاهدات مطروح منها الرقم واحد إذا كنا نتعامل مع العينة ونريد إيجاد أفضل التقدير، وبشكل عام هذة الخطوة تعطي تباين المجتمع أو أفضل تقدير له،
٨. للحصول على الانحراف المعياري نوجد الجذر التربيعي لقيمة ناتج العملية الحسابية رقم (٧).

وعلى الرغم من أن الخطوات السابقة قد توحي بأن العملية بسيطة ولكن إذا كان عدد المشاهدات كبيراً فإن مقدار الجهد المطلوب لحساب الخطوات السابقة سيكون كبيراً فضلاً عن الخطأ الذي قد يقع فيه الباحث إذا تم الحساب بشكل يدوي. ومن ثم فإن استخدام برنامج SPSS أو أي برنامج احصائي آخر، هو البديل لتجنب مثل هذه الأخطاء وذلك من خلال اختيار المقياس المناسب وتحديد بيانات المتغيرات قيد الدراسة داخل البرنامج. ومنذ ظهور البرنامج عام ١٩٦٨م ومع زيادة ابتكار مقاييس إحصائية عديدة لمعالجة المشكلات الجديدة التي تظهر في العلوم فكان من المناسب مواكبة البرنامج لهذا التقدم ليشمل معظم العلوم فضلاً على تصميم اختبار الاستمارة الإحصائية أيضاً، مما جعل البرنامج أكثر انتشاراً واستخداماً. ومن ناحية توافق البرنامج لبرامج التشغيل فقد كان

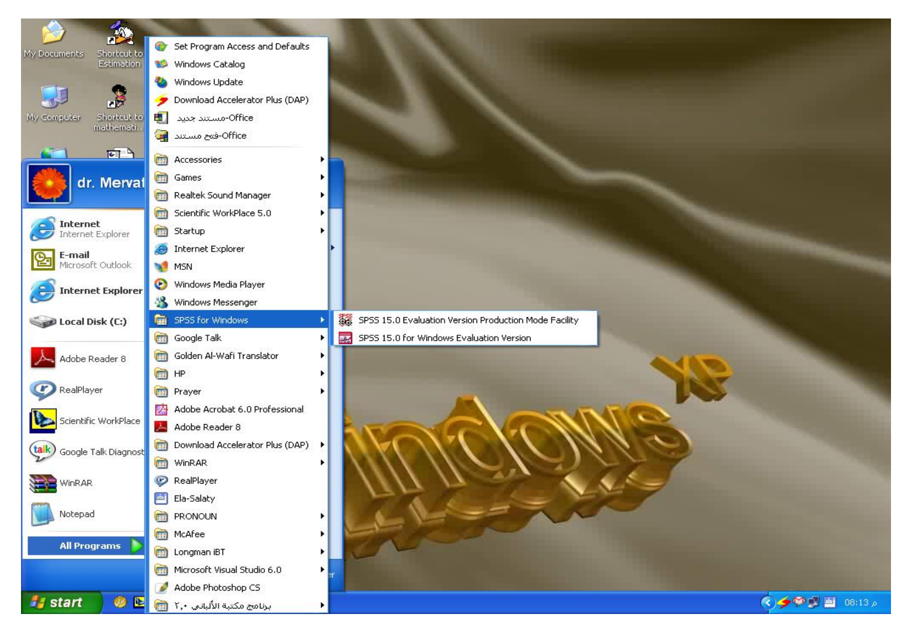
التطور في ذلك مناسباً جداً حيث كان البرنامج يعمل تحت نظام التشغيل MS-DOS وتم تطويره ليعمل في بيئة نظام التشغيل WINDOWS في عام ١٩٩٣م متلافياً بذلك الصعوبات التي كانت تواجه مستخدمي هذا النظام في بيئة MS-DOS. وقد توالت الاصدارت لهذا النظام التي كان أخرها الإصدار الخامس عشر حيث يوفر هذا النظام مجالا واسعاً للتحليلات الإحصائية وإعداد المخططات البيانية لتلبية حاجة المختصين والمهتمين في مجال الإحصاء. كما يوفر إمكانية تناقل البيانات مع قواعد البيانات وبرامج EXCEL و LOTUS وغيرها من البرامج الأخرى وبالتالي فإن هذه الإمكانيات مناسبة جداً لموضوعات الكتاب.

الدخول للبرنامج (Getting Start):

يوجد لدينا طريقتين للدخول إلى البرنامج:

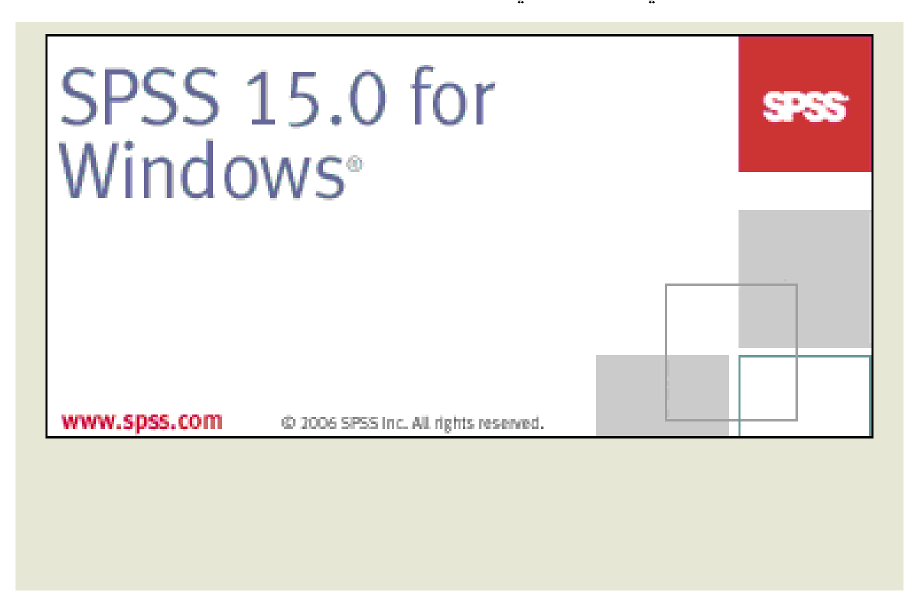
- **الطريقة الأولى:**

بعد تحميل البرنامج ننقر على Start في شريط المهام ومن ثم النقر على Programs ثم اختيار SPSS كما في الشكل (١) التالي.



الشكل(١)

نبدأ الدخول إلى البرنامج وذلك بظهور الشاشة الافتتاحية للبرنامج وهو إعلان عن الحزمة لمدة ثوان، كما في الشكل التالي:



الشكل (٢)

بعد ظهور الشاشة الافتتاحية تبدأ أول شاشة من شاشات الحزمة في الظهور.

⌖ **الطريقة الثانية:**

الدخول بهذه الطريقة يتم عن طريق أيقونة مختصرة (Short Cut Icon) أعدت ووضعت على سطح المكتب عن طريق المستخدم ويتم إعداد هذه الأيقونة من خلال النقر على اسم البرنامج في الشكل (١) بمفتاح الفارة الأيمن واختيار Send to ثم اختيار Created Shortcut، ثم يظهر علي سطح المكتب أيقونة البرنامج وبالنقر مرتين عليها فندخل للبرنامج مباشرة. وهذه الأيقونة تكون كما بالشكل الأتي:



الشكل (٣)

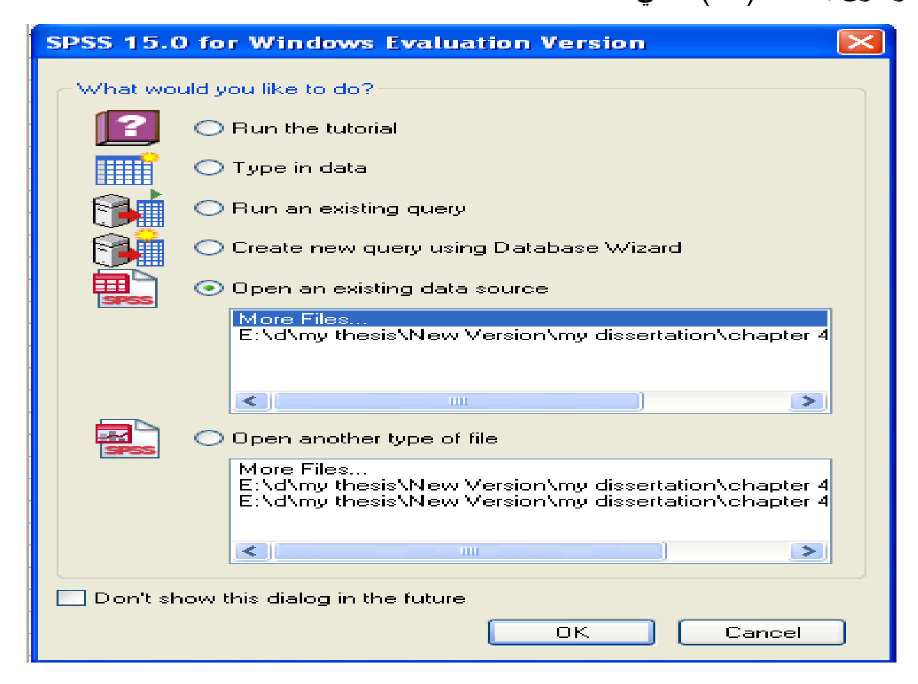
ملاحظة: في هذا الإصدار وبعد ظهور الشاشة الافتتاحية تصدر شاشة اختيارية بعنوان:

What Would You Like To Do?

هذه الشاشة اختيارية وبها عدة اختيارات منها هل يطلب المستخدم تشغيل البرنامج التعليمي، إدخال البيانات، فتح ملف، ...الخ، يمكن للمستخدم اختيار ما يريد تنفيذه وذلك بالنقر على الاختيار المطلوب، ثم النقر على Ok أو النقر على Cancel للانتقال إلى الشاشة التالية. ويمكن للمستخدم إعطاء أمر للبرنامج بعدم إظهار هذه الشاشة مرة أخرى وذلك بالنقر على العبارة:

(Do Not Show This Dialog In The Future)

وتكون بالشكل (٤) التالي:



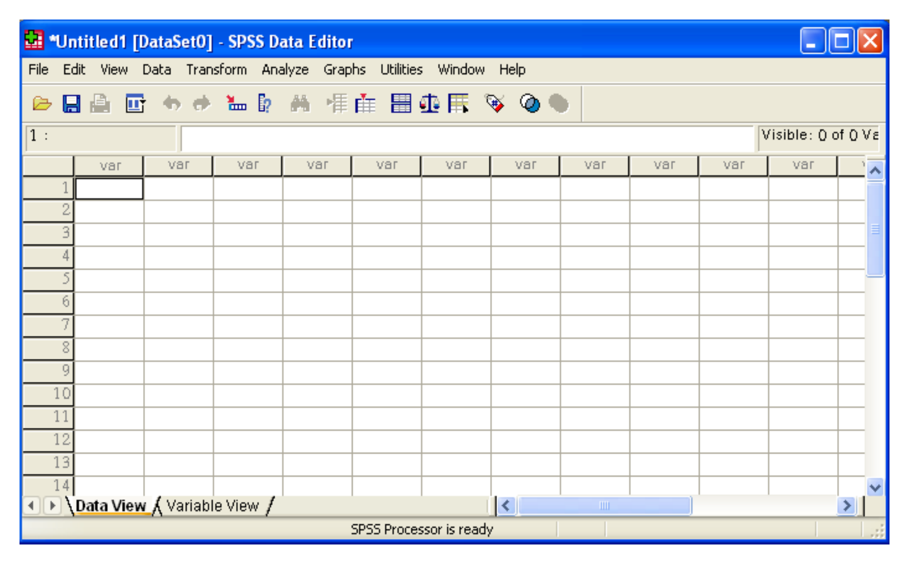
الشكل (٤)

بيئة الحزمة (The SPSS Environment):

تعمل الحزمة SPSS تحت مجموعة من النوافذ تمثل في مجموعها منظومة للتعامل مع البرامج، وهي كالتالي:

- **نافذة محرر البيانات (Data Editor):**

تعرض هذه النافذة محتويات ملف معين من البيانات حيث يمكن انشاء ملف جديد أو تحرير ملف موجود وإن هذه النافذة تفتح تلقائيا عند بدء تشغيل البرنامج كما بالشكل (٥) التالي:



الشكل (٥)

مع العلم أن بيانات هذه الشاشة قابله للتعديل والحفظ والطباعة للحصول على الأشكال البيانية والجداول بالإضافة إلى العرض والتحليل الإحصائي للحصول على المقاييس والاختبارات الإحصائية المتاحة.

- **نافذة شاشة المخرجات (Viewer Window):**

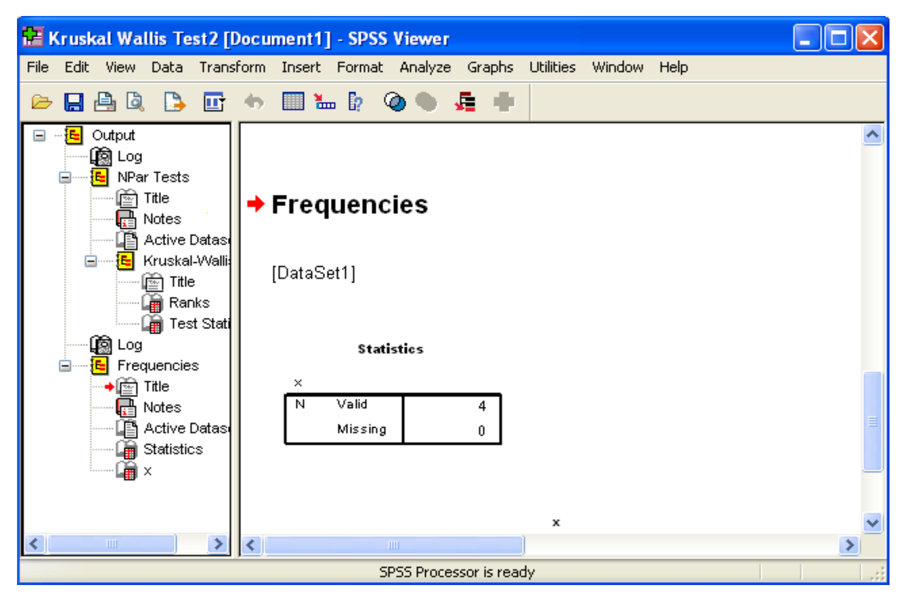
يعرض في هذه الشاشة كل النتائج الإحصائية (جداول إحصائية، مقاييس إحصائية، اختبارات إحصائية، أشكال بيانية،... الخ) وهي قابلة للتعديل والحفظ وهذه التعديلات تشمل الجداول والأشكال البيانية والنصوص، ... الخ.

تنقسم هذه الشاشة إلى جزئيين:

جز ء على اليسار يحتوي على فهرس تفصيلي لمحتويات الشاشة وجز ء على اليمين لعرض محتويات الشاشة من جداول وأشكال بيانية ونتائج وتحليلات إحصائية، ... الخ .

كما تحتوي الشاشة على شريط للأدوات وشريط للأوامر كما في شاشة محرر البيانات.

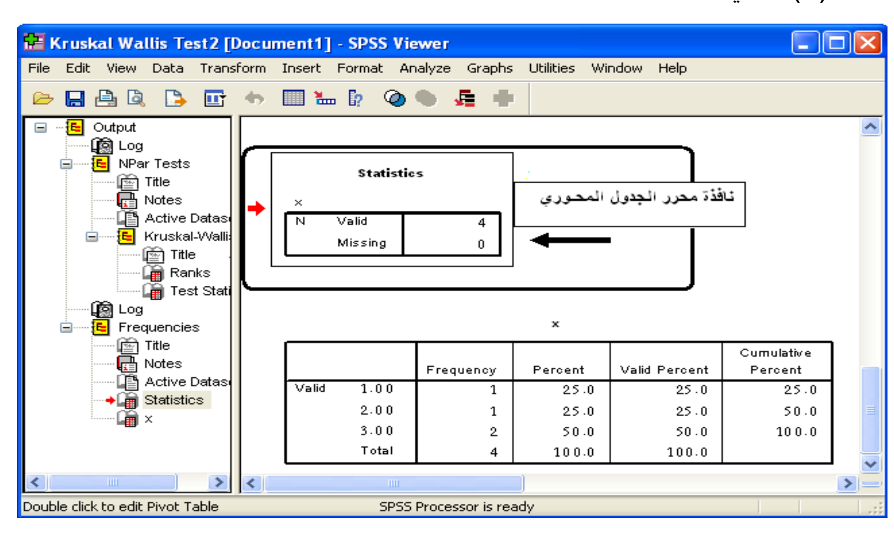
انظر الشكل (٦) التالي:



الشكل(٦)

▪ **نافذة محرر الجدول المحوري (Pivot Table Editor):**

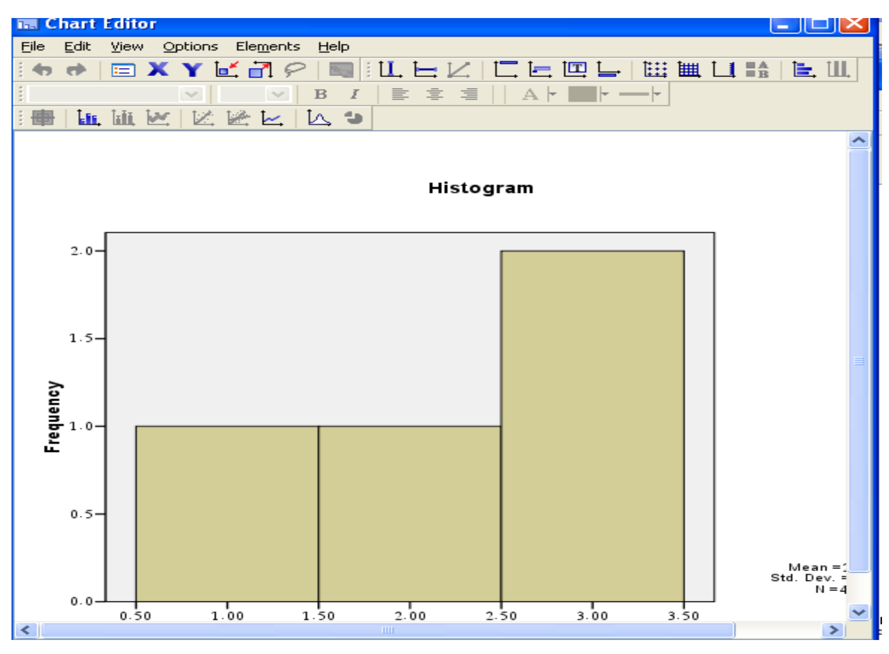
يمكننا اجراء تعديل على الجداول في صفحة النتائج وذلك بالنقر على الزر الأيمن بالفأرة على الشكل المراد تعديله فيتم تحديد الجدول عندئذ يمكن تعديل أو إلغاء صفوف وأعمدة من الجداول وكذلك تغيير الألوان وإنشاء جداول متعددة الاتجاهات، انظر الشكل (٧) التالي.



الشكل(٧)

- **نافذة محرر الأشكال (Chart Editor) :**

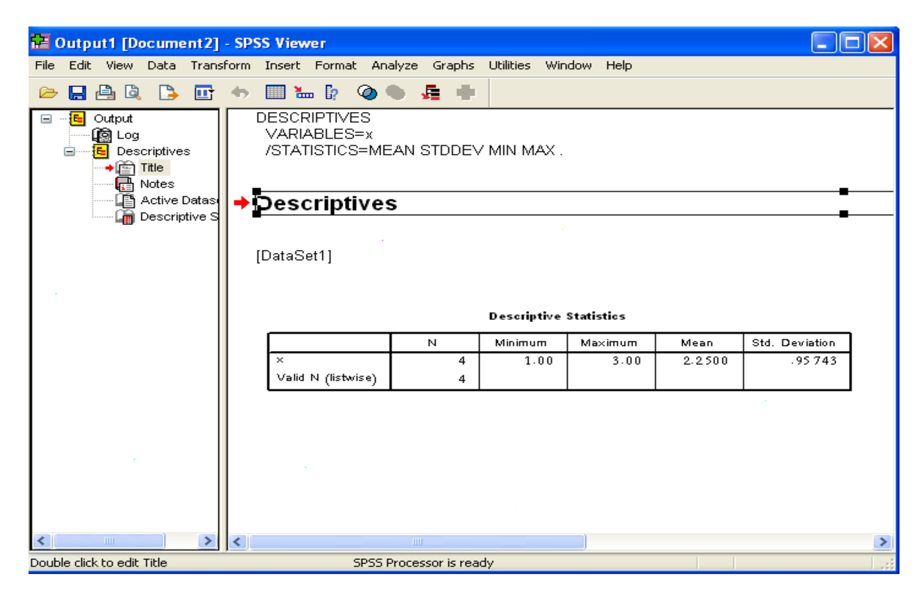
يمكن اجراء تعديل على الأشكال البيانية و الرسومات بواسطة محرر الأشكال مثل تعديل (اللون وحجم الخط ونوعه وتغيير المحاور) عند فتح صفحة النتائج والنقر المزدوج بالمفتاح الأيمن للفأرة على الشكل المراد تعديله سيؤدي إلى فتح شاشة المحرر وظهور شريط الأدوات الخاص بهذه الشاشة ويمكن العودة إلى شاشة النتائج بالنقر المزدوج على الشكل المعدل، انظر الشكل (٨) التالي:



الشكل(٨)

- **نافذة محرر النص (Text Output Editor):**

يقوم محرر النص بتعديل النصوص للأشكال البيانية مثل (العنوان أو ملحق الشكل أو الجدول) وذلك بالنقر المزدوج بالمفتاح الأيسر للفأرة على النص المراد تعديله فيتم تحديده بخطوط للتعديل علما أن التعديل يشمل اللون ونوع الخط وغيره، انظر الشكل (٩) التالي:



الشكل (٩)

- **نافذة محرر القواعد (Syntax Editor) :**

هذه الشاشة تحتوي على الأوامر والتعليمات التي تم اختيارها لبعض عمليات العرض أو التحليل المكتوبة بلغة SPSS ويمكن تعديلها وحفظها وإعادة تشغيلها. كما يوجد في قائمة الأوامر الأمر Run الذي يسمح بتشغيل الأوامر المعروضة مرة أخرى.

- **نافذة محرر الدور (Script Editor):**

هذه الشاشة لها دور مهم في سهولة التعامل مع الحزمة SPSS حيث يمكن عن طريقها إنشاء أو تعديل الخطوط الأساسية للحزمة الجاهزة.

أنواع الملفات في البرنامج:

يتوفر في برنامج SPSS عدة أنواع من الملفات منها ما يلي:

- **ملفات البيانات Data Files:**

تتكون هذه الملفات باستخدام محرر البيانات Data Editor وهي تحتوي على البيانات التي تستخدم في التحليل الإحصائي ويكون لهذا النوع من الملفات الامتداد SAV.

- **ملفات المخرجات الإحصائية Output Files:**

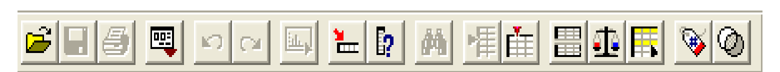
وهي ملفات تحتوي على مخرجات التحليل الإحصائي أو المخططات وتكون ذات امتداد SPS.

- **ملفات التعليمات Syntax:**

وهي الملفات التي تحوي الإجراءات الإحصائية التي تخزن على شكل أوامر وتكون ذات امتداد SPS.

شريط الأدوات (SPSS Toolbar):

هذا الشريط يحتوي على اختصار لبعض العمليات في صورة أيقونـات رسـومية، كل رسم يمثل عملية معينة مثل (فتح ملف، حفظ ملف، ... الخ) الشكل التـالي يوضـح شريط الأدوات في صفحة المحرر.



الجدول التالي يوضح وصف مختصر لمحتويات شريط الادوات:

جدول (١)

الوظيفة	العنوان	الأيقونة
فتح ملف	File Open	
حفظ ملف	File Save	
طباعة ملف	File print	
إظهار آخر مجموعة من الإجراءات التي تم استخدامها	Dialog recall	
التراجع عن آخر تغيير	Undo	
إعادة اجراء التغيير	Redo	
الانتقال إلى الرسم البياني	Go to chart	
الانتقال إلى الحالة	Go to case	
عرض معلومات عن المتغيرات	Variables	
البحث عن حالة ضمن متغير	Find	

إضافة حالة	Insert case	
إضافة متغير	Insert variable	
تجزئة ملف	Split file	
إعطاء أوزان لبعض المتغيرات	Weight cases	
اختيار حالات	Select cases	
إظهار أو أخفاء عناوين القيم	Value labels	
استخدم مجموعات	Use sets	

وهناك عدة طرق مختلفة للحصول علي المساعدة:

- قائمة Help : ومما سبق يتضح أن هذة القائمة في شريط القوائم Menu bar لكل نافذة من نوافذ SPSS وتتضمن Topics ومن خلالها يتم توفير ثلاثة أنواع من المساعدة بواسطة Contents، Index و Find كما تتضمن Tutorial التي توفر مدخلاً تعليمياً إلي برنامج SPSS.
- زر المساعدة في صندوق الحوار Dialog box help button هذا الزر موجود في أغلب صناديق الحوار لبرنامج SPSS ومن خلاله يمكن الحصول على معلومات عامة عن الموضوع المتعلق بصندوق الحوار.
- المساعدة الموضعية في صندوق الحوار Dialog box context menu help : يمكن الحصول على مساعدة عن أي نص يرد في صندوق الحوار بنقر ذلك النص بزر الماوس الأيمن Right-Click لعرض وصف عن ذلك النص.
- المساعدة المرتبطة بسياق الجدول المحوري Pivot table context menu help : يمكن الحصول علي هذة المساعدة بنقر الجدول المحوري مرتين بمفتاح الماوس الأيسر لتفعيله ثم ضغط عنوان Label الصف أو العمود (في حالة أنه

يمثل مؤشراً تم احتسابه من خلال البرنامج) بنقر مفتاح الماوس الأيمن في الجدول المحوري في شاشة SPSS Viewer (علما أن معظم مخرجات SPSS هي جداول محورية) ثم اختيار What's this من القائمة الموضوعية لعرض تعريف عن محتويات الصف أو العمود.

- مرشد النتائج Result Coach: يمكن الحصول على هذا النوع من المساعدة بنقر الجدول المحوري مرتين بمفتاح الماوس الأيسر لتفعيلة ثم ضغط الجدول بمفتاح الماوس الأيمن واختيار Result Coach من القائمة الموضعية لعرض تفسير إحصائي مبسط للنتائج الإحصائية المضمنة في الجدول من خلال عدة نوافذ متسلسلة.
- المرشد Tutorial: يمكن الحصول على هذه المساعدة باختيار Tutorial من قائمة help في أي نافذة للوصول إلى مدخل تعليمي مباشر باستعمال عدد من النوافذ التعليمية المتسلسلة.

المربع الحواري (Dialogue Box):

يتيح هذا الإصدار من البرنامج اختيار المتغيرات التي نرغب في اجراء التحليلات الإحصائية عليها عن طريق المربعات الحوارية، كما أنه يعتبر بديلا عن كتابة أوامر البرمجة المعقدة، عند طلب أي أمر من الحزمة SPSS علماً بأن البرنامج يظهر مربعاً حوارياً يحتوي على مجموعة من الخيارات كل خيار يمثل بزر (أمر) وللباحث اختيار ما يناسبه ليضعها موضع التنفيذ وبسرعة فائقة. يتكون مربع الحوار في برنامج SPSS من العناصر التالية:

- قائمة متغيرات المصدر Source Variables List: وتشمل كافة المتغيرات الموجودة في الملف الحالي ذات الأنواع المسموحة الاستخدام للأسلوب الإحصائي المختار.
- قائمة متغيرات الهدف Target Variables List: واحدة من القوائم التي تتضمن أسماء المتغيرات المختارة للتحليل الإحصائي عليها.
- أزرار الأوامر Command Pushbuttons: و هذه المفاتيح تقوم بإعلام البرنامج لتنفيذ عمل معين مثلا تمشية البرنامج أو الحصول على مساعدة.

الفصل الأول

تجهيز ملفات البيانات

Data File Preparation

استعرضنا في الفصل السابق بعض الخطوات العلمية للتحليل الإحصائي والبحث العلمي بشكل عام. في هذا الفصل نناقش طريقة تعريف المتغيرات وإدخال البيانات محل الدراسة للبرنامج تمهيداً لإجراء التحليل الإحصائي عليها.

المتغير والحالة (Variable and Case):

المثال التالي يعطي فكرة مبسطة عن المتغير والحاله.

مثال (١):

البيانات التالية تمثل قيد مجموعة معينة من الأشخاص في اختبار معين.

جدول (١)

Name	ID	Gender	Age	Birth date	Income
Mohammed	1	1	38	1/1/1970	1000
Mervat	2	2	27	20/8/1980	500
Ahmad	3	1	48	15/3/1960	2000
Heba	4	2	22	15/4/1985	300
Nabil	5	1	18	1/1/1990	200
Noha	6	2	58	20/7/1949	3000

جدول بيانات التحليل: كل عمود في المثال يمثل متغير من المتغيرات، والمتغيرات يمكن تقسيمها كما يلي:

- المتغير الأول: متغير الاسم ID وهو متغير رمزي.
- المتغير الثاني: متغير الجنس Gender (الرقم ١ يمثل الذكور والرقم ٢ يمثل الإناث).
- المتغير الثالث: متغير العمر Age وهو متغير رمزي أيضا.
- المتغير الرابع: متغير تاريخ الميلاد Birth date وهو متغير تاريخ .
- المتغير الخامس: متغير مستوي الدخل Income وهو متغير رقمي.

أنواع البيانات في البرنامج:

- **بيانات كمية**: يستعمل للبيانات العددية (القابلة للقياس الكمي) مثل بيان الدخل ودراجاتك في الامتحان وهكذا.
- **بيانات مكودة**: يستعمل لقياس المتغيرات الاسمية والتي لا يمكن ترتيبها تصاعديا أو تنازليا كما لا يمكن اجراء العمليات الحسابية على هذا النوع من المتغيرات مثل / تقسيم المجتمع إلى ذكور وإناث و الأسماء أيضا يمكننا اعتبارها مقياس اسمي.
- **بيانات تاريخ**: وهي البيانات التي نرغب بتوضيح بدايتها ، نهايتها، أو فترتها بالتاريخ الزمني.
- **بيانات عملات**: وهي البيانات التي نرغب بتوضيح العملات المستخدمة في حسابها.
- البيانات النصية: وهي البيانات التي نرغب في كتابتها بشكل نصي في البرنامج مثل اسم المتغيرات.

تعريف الاكواد (Codes):

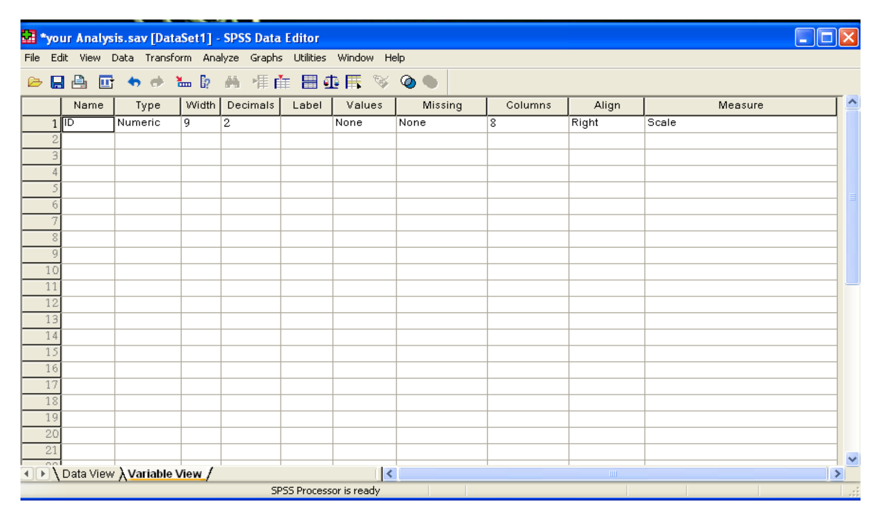
نقصد بالأكواد تحويل المتغير اللفظي إلى رقمي ويطلق على الرقم الذي يمثل قيمة المتغير اللفظي كود أو رمز. يمكن التمييز بين المتغيرات سواء كانت لفظية أو رقمية بناء على ما يسمى وحدة القياس .

تعريف المتغيرات (Defining Variables):

تعريف المتغير في برنامج SPSS يعنى أنه يجب تحديد ما يلي:

- اختيار اسم المتغير Variables Names.
- اختيار مميز المتغير Variable label.
- اختيار قيمة المميز إن وجد Value label.
- تعريف القيم المفقودة Missing Values.
- تحديد نوع المتغير Variable Type.
- تحديد شكل عرض البيانات في الأعمدة Column format .

حيث نقوم بتعريف المتغير من شاشة Variable View الظاهرة في الشكل التالي:



شكل (١)

وسنقوم بشرح كل عمود من أعمدة الشاشة السابقة لكون تلك الشاشة شاشة تعريف متغيرات تحليلك الإحصائي.

- **اختيار اسم المتغير:**

تبدأ عملية تعريف المتغير بتحديد اسما للمتغير في العمود الأول من أعمدة Variable View والظاهر في الشكل وذلك بالنقر عليه نقراً مزدوجاً بزر الماوس الأيمن وكتابة اسم المتغير بعد ذلك باستخدام لوحة المفاتيح الخاصة بجهازك.

توجد عدة شروط يجب أن تؤخذ في الاعتبار عند تحديد أسم المتغير هي:

- ✓ يجب أن لا يزيد عن ٦٤ حرفاً (for version 15 or 12) أو ثمانية حروف أو أرقام لنسخ السابقة.
- ✓ أن يبدأ الاسم بحرف (ليس رقماً) ولا يكون به فراغات .
- ✓ لا يسمح باستخدام الحروف التالية كجزء من اسم المتغير (. / ! / ؟ / ، / *).
- ✓ لابد أن يكون الاسم وحيد في نفس الملف بمعنى أن كل متغير يأخذ اسم لا يأخذه متغيراً آخر.
- ✓ لا يمكن استخدام الكلمات الآتية كاسم للمتغير With – Ge – Not – Ne – Eq – To – Le -And – Gt – Or - It by – all لأن هذه الكلمات لها استخدامات معينة داخل الحزمة SPSS .

✓ يمكن كتابة الاسم باللغة الإنجليزية سواء بحروف كبيرة أو صغيرة.

▪ **تعريف نوع المتغير** (Kind of Variable):

هذا هو العنصر الثاني من تعريف المتغيرات محل اهتمامك في التحليل الإحصائي إذ وهو نوع المتغير وهو العمود الثاني من Variable View وفي الحقيقة يعتبر هذا العنصر في منتهى الأهمية بالنسبة للتحاليل الإحصائية حيث تختلف أنواع المقاييس الإحصائية المستخدمة على حسب نوع المتغيرات محل الدارسة. ومعظم الاختيارات تكون رقمي "numeric" هذا يعني أن المتغير يأخذ رقماً لاحظ الآتي أن كلمة رقماً لا تعني بتأكيد قيمة عددية يمكننا من خلالها إجراء العمليات الحسابية المعتادة ونجد أن هناك خيار آخر وهو "string" وهو خاص بالمتغيرات التي تأخذ نصوص في التحليل الإحصائي والجدول الآتي يوضح أنواع البيانات المتاحة في برنامج SPSS:

جدول (٢)

نوع المتغير	مثال
Numeric	1000.36
Comma	000.005،1
Scientific	1*e3
Dollar	000.00،$1
String	Mervat

ونلاحظ أن البرنامج يضع بشكل تلقائي نوع المتغير Numeric ويمكننا تغيرها للأسباب الآتية:

- أننا نرغب في توضيح فواصل عشرية كثيرة أو قليلة جداً.
- أن أرقام المتغير كبيرة جداً مثل 12456789123 فهنا على سبيل المثال نستطيع اختيار الخيار comma لكي نستطيع أن نقرأ الرقم فيصبح الرقم 12،456،789،123، أو نضعها باستخدام scientific فتصبح 12.45*E9 فنستطيع قراءتها بسرعة على أنها ١٢ بليون (E3 هي ترمز إلي الآلاف، E6 ترمز إلى المليون ، و E9 ترمز إلى البليون).

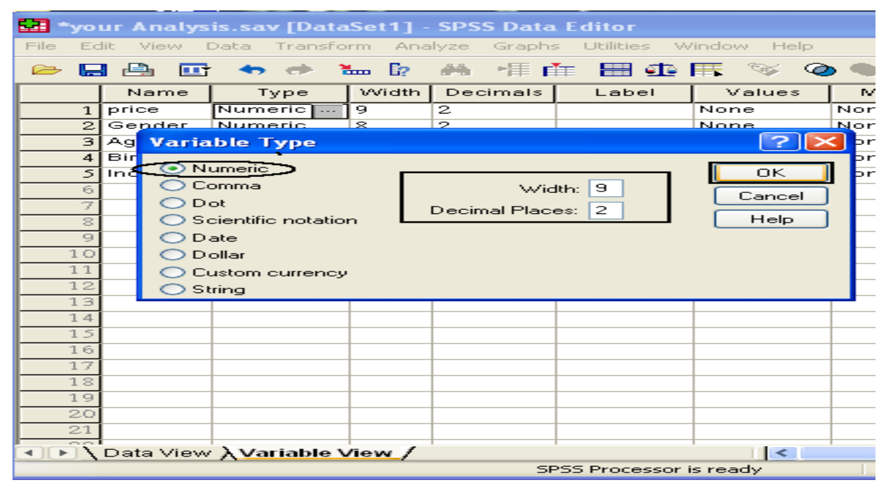
⌖ في حالة ما تكون المتغير هي عملات ونرغب في توضيح ذلك فإننا نضع نوع المتغير Dollar.

والآن نستعرض بعض الأمثلة لتوضح كيف يمكن تغير نوع المتغير في البرنامج:

مثال (٢) بيانات كمية:

لتغير نوع البيانات إلى رقمي نتبع الخطوات الآتية:

نختار Variable view ⇦ وبعد ذلك عند المتغير الذي نريد أن نغير نوعه نختار العمود الثاني عنده والذي يحمل عنوان Type ونلاحظ أن بجوار نوع المتغير توجد نقاط على شكل " ... " نقوم بنقر عليها فيظهر الشكل (٢) التالي:



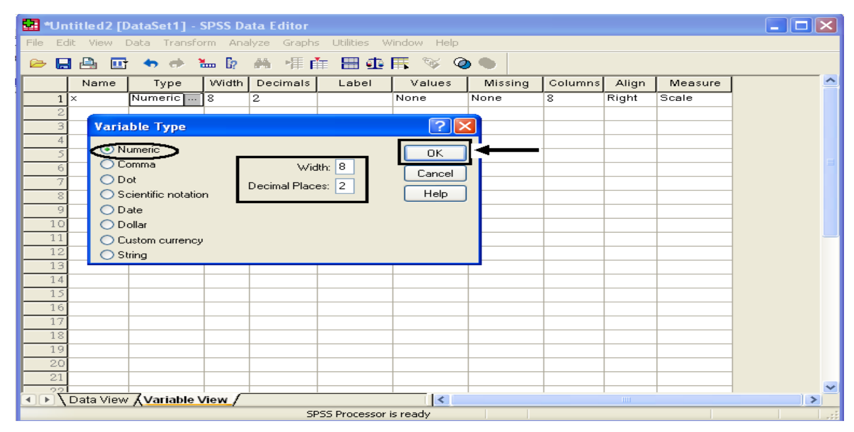
شكل (٢)

نختار Numeric وهي أول خيار ونستطيع من ذلك عرض عمود البيانات لهذا المتغير بحيث يظهر عدد من خانات الأرقام أكبر وهنا نختار 9 Width وعدد الخانات العشرية 2 بحيث إذا كان المتغير يأخذ الرقم الآتي 123456789.659 فيظهر في البرنامج بالشكل الآتي 123456789.66 فيقوم البرنامج بتقريب الخانات العشرية إلي أقرب رقمين فقط . وبعد ذلك ننقر على Ok لتأكيد الاختيار ونلاحظ أننا بعد النقر على Ok يتغير قيمة العمود الثالث الذي يحمل عنوان Width إلي 9 وهو عدد خانات الرقم الأساسية، ويتغير أيضا قيمة العمود ٤ والذي يحمل عنوان Decimals إلى 2 وهو عدد الخانات العشرية المتاحة .

مثال (٣): البيانات المكودة.

البيانات المكودة هي البيانات الاسمية التي نعطي لها رقماً يعبر عنها في البرنامج وهي من أهم أنواع البيانات في البرنامج من هذه البيانات بيان النوع أو الجنس بيان الحالة الزوجية ففي بيان الجنس نرغب بتعبير عن حالة النوع" ذكر" برقم" ١" وحالة النوع أنثى برقم" 2" ونلاحظ أن تلك الأرقام هي أكواد وليست قيماً يمكننا إجراء العمليات الحسابية عليها. ولتعريف البرنامج بالبيانات المكود نتبع الخطوات التالية:

نختار Variable view ⇦ وبعد ذلك عند المتغير الذي نريد أن نغير نوعه نختار العمود الثاني عنده والذي يحمل عنوان Type نقوم بنقر عليها فيظهر الشكل (٣) التالي:



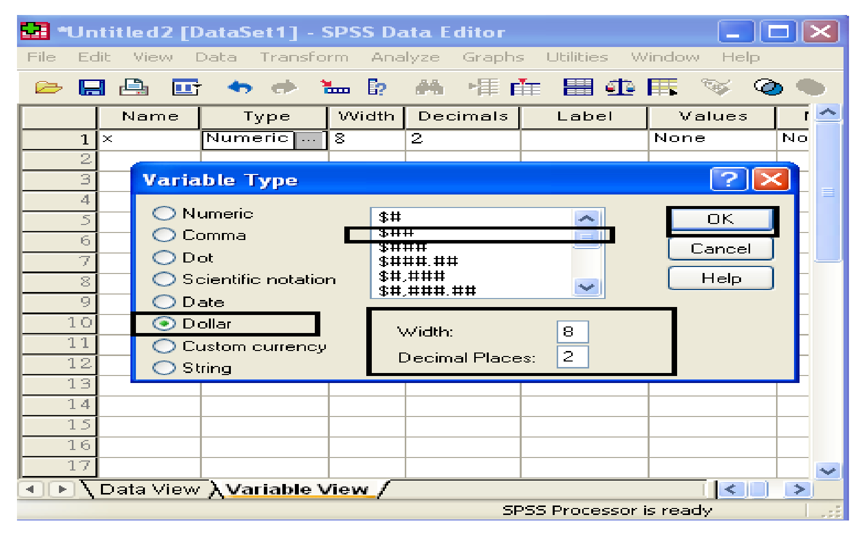
شكل (٣)

نختار Numeric وهو أول خيار ونستطيع من ذلك عرض عمود البيانات لهذا المتغير بحيث يظهر عدد من خانات الأرقام أكبر وهنا نختار Width 2 وعدد الخانات العشرية 0 حيث أننا لدينا خانتين فقط وليس لدينا أرقام عشرية وبعد ذلك النقر على Ok لتأكيد الاختيار.

مثال (٤): بيانات العملات

في حالة البيانات التي تعبر عن عملات معينة نريد توضيح أن ذلك البيان هو بيان يتبع عملة معينة مثل بيان الدخل فإننا نعرف ذلك البيان إلى البرنامج على طريق أختيار view Variable ⇦ وبعد ذلك عند المتغير الذي نريد أن نغير نوعه نختار العمود الثاني عنده والذي يحمل عنوان Type نقوم بنقر عليها ومن ثم نقوم بتحديد الخيار

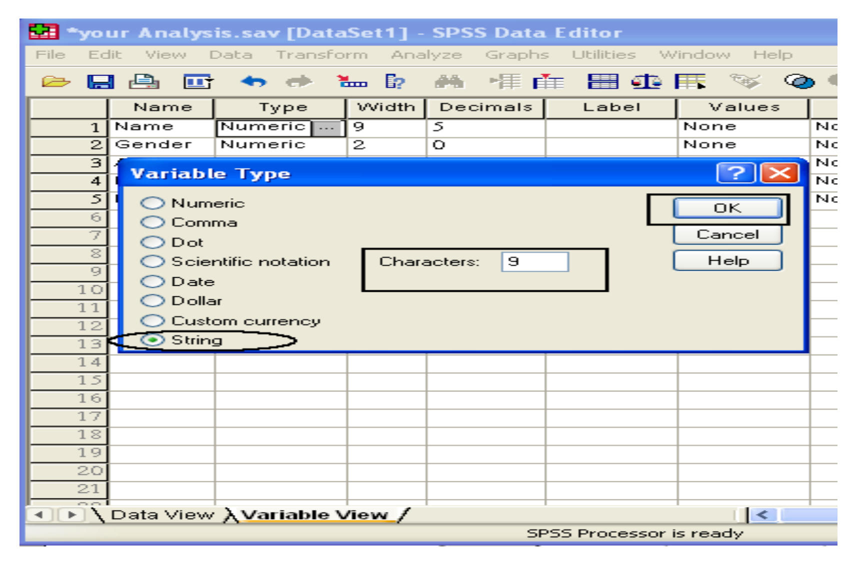
Dollar يلي ذلك تحديد عدد خانات بيان العملة المرغوب توضيحها وبعد ذلك نقوم بنقر على ok لتأكيد الاختيار فيظهر الشكل (٤) التالي:



شكل (٤)

مثال (٥): بيانات نصية

في حالة البيانات النصية إذا أردنا كتابة المتغيرات النصية في البرنامج بشكل نصي فإننا نعرف ذلك البيان إلى البرنامج على طريق الخطوات التالية: نختار Variable view ⇦ وبعد ذلك عند المتغير الذي نريد أن نغير نوعه نختار العمود الثاني عنده والذي يحمل عنوان Type نقوم بنقر عليها ومن ثم نقوم بتحديد الخيار string يلي ذلك تحديد عدد خانات البيان النصي المرغوب وضعه وهو أقصى عدد أحرف يمكن إضافتها وبعد ذلك نقوم بنقر على ok لتأكيد الاختيار فيظهر الشكل (٥) التالي:

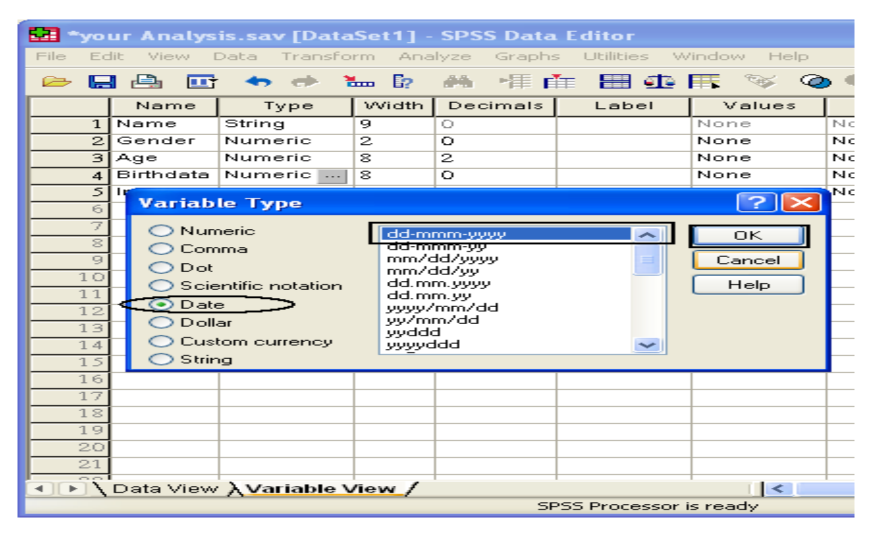


شكل (٥)

مثال (٦): بيانات التاريخ

إذا أردنا كتابة متغير بيان تاريخ الميلاد في البرنامج فإننا نعرف ذلك البيان إلى البرنامج على طريق الخطوات التالية:

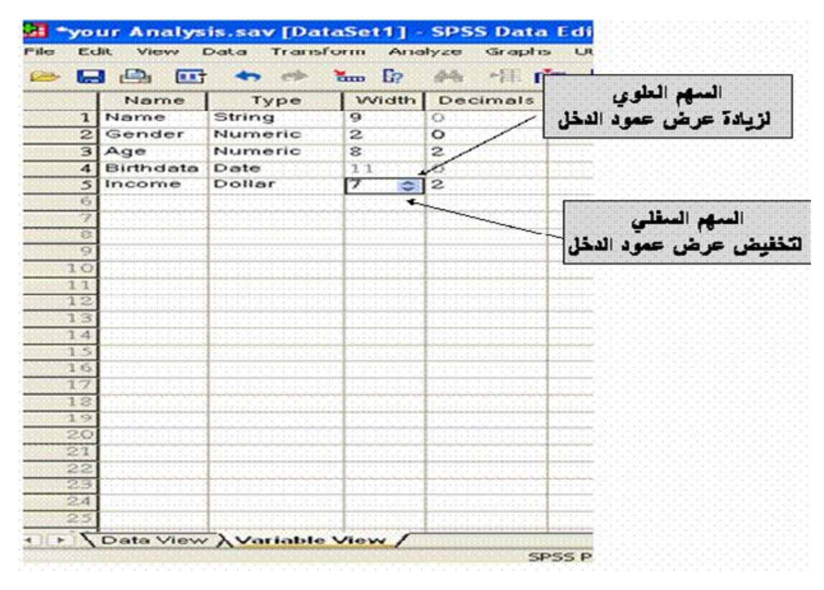
نختار Variable view ⇦ وبعد ذلك عند المتغير الذي نريد أن نغير نوعه نختار العمود الثاني عنده والذي يحمل عنوان Type نقوم بالنقر عليها ومن ثم نقوم بتحديد الخيار data يلي ذلك تحديد نمط التاريخ، فإذا افترضنا أننا نتبع النظام البريطاني في كتابة التاريخ حيث يكتب اليوم في البداية ومن ثم كتابة الشهور وفي النهاية كتابة السنوات بكامل الشكل فإننا ننقر على النمط الأول لتاريخ وبعد ذلك نقوم بنقر على ok لتأكيد الاختيار فيظهر الشكل (٦) التالي:



شكل (٦)

عرض المتغير (Width Variable):

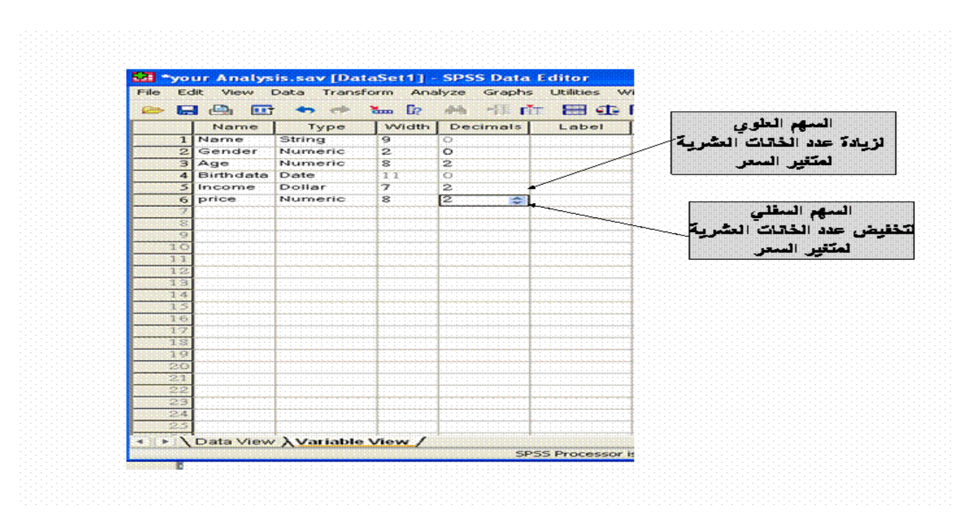
هذا هو العنصر الثالث من تعريف المتغيرات محل اهتمامك في التحليل الإحصائي إذ وهو عرض المتغير وهو العمود الثالث من Variable View وفي الحقيقة يمكننا من تغير عرض المتغير مباشرة عندما نقوم بتعريف نوع المتغير أو بنقر مباشرة على الخلية الموجودة في عمود عرض المتغير وصف المتغير المراد تغير عرضه فيظهر سهم إلي أعلى وسهم إلي أسفل وعند النقر على السهم العلوي يقوم بزيادة عرض المتغير مقدار وحدة في كل مرة ضغط وهكذا بالنسبة للأسهم السفلي حيث يقوم بتخفيض عرض المتغير بمقدار وحدة واحدة في كل مرة ضغط والشكل (٧) يوضح هذه العملية:



شكل (٧)

عدد الخانات العشرية للمتغير (Decimals of Variable):

هذا هو العنصر الرابع من تعريف المتغيرات محل اهتمامك في التحليل الإحصائي إذ وهو عدد الخانات العشرية للمتغير وهو العمود الرابع من Variable View حيث يمكننا من تغيير عدد الخانات العشرية للمتغير مباشرة عندما نقوم بتعريف نوع المتغير أو بنقر مباشرة على الخلية الموجودة في عمود عدد الخانات العشرية للمتغير (Decimals) وصف المتغير المراد تغيير عدد خانته العشرية فيظهر سهم إلى أعلى وسهم إلى أسفل وعند النقر على السهم العلوي يقوم بزيادة عدد الخانات العشرية للمتغير بمقدار وحدة في كل مرة ضغط وهكذا بالنسبة للسهم السفلي حيث يقوم بتخفيض عدد الخانات العشرية له أيضا بمقدار وحدة واحدة في كل مرة ضغط والشكل (٨) يوضح هذة العملية:



شكل (٨)

مميز المتغير (Variable label):

يمكن أن يأخذ المتغير عنوانا يصل عدد رموزه إلى ٢٥٦ رمز يستعمل لوصف المتغير مع إمكانية وضع فواصل ورموز خاصة بعكس تماماً خانة اسم المتغير فيمكنا من تعريف متغير الاسم على أنه اسم العميل Name of the agent كما بالشكل (٩) التالي:

*your Analysis.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values
1	Name	String	9	0	Name of our agent	None
2	Gender	Numeric	12	1		None
3	Age	Numeric	8	6		None
4	Birthdata	Date	11	0		None
5	Income	Dollar	6	4		None
6	price	Numeric	8	6		None

Data View / Variable View

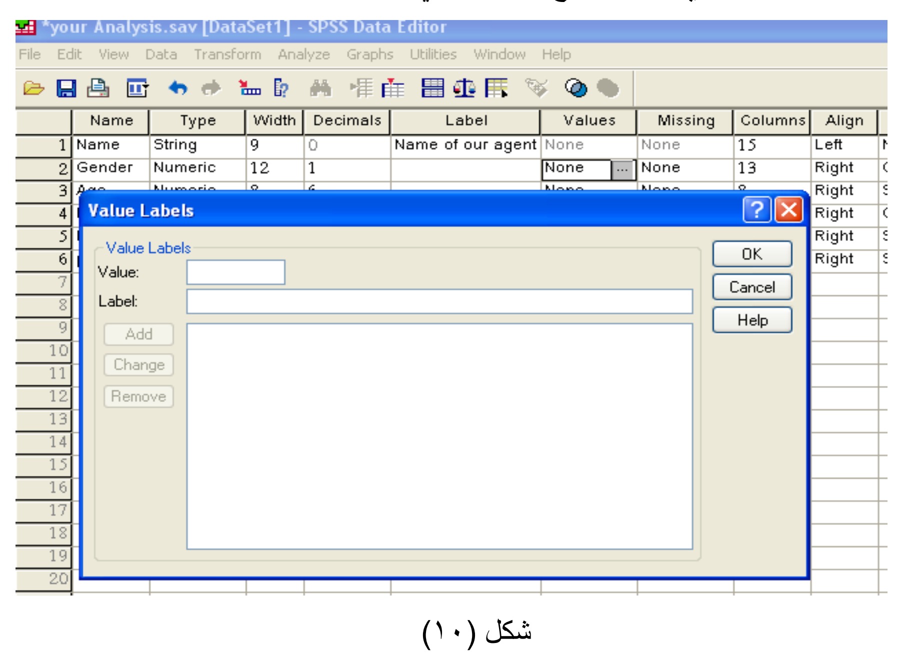
SPSS Processor is ready

شكل (٢-٩)

قيمة المميز (Value label) :

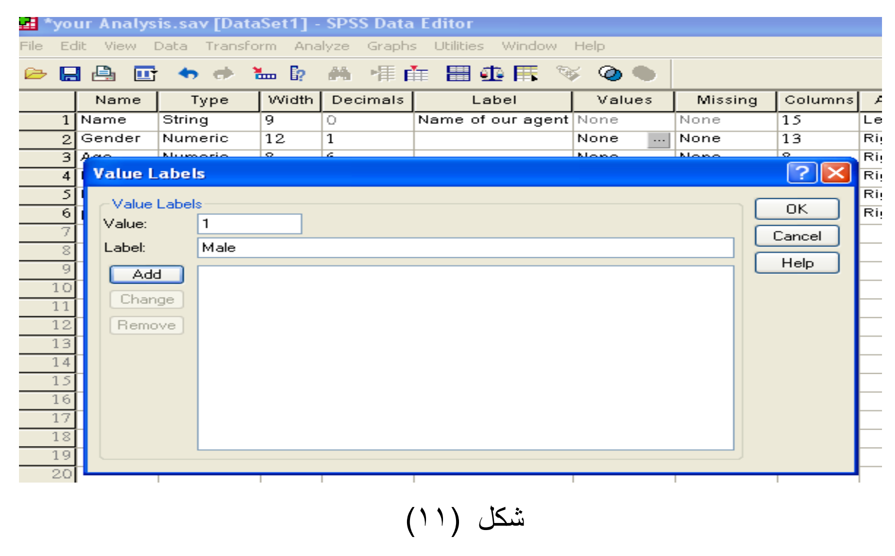
أحيانا تبرز الحاجة إلى تعيين عنوان للقيمة كون المتغير يستعمل قيما عددية للتعبير عن قيم غير عددية مثلا / متغير Gender يستعمل الرقم ١ للتعبير عن الذكور والرقم ٢ للتعبير عن الإناث ونلاحظ أن هذا التعريف معرف لجميع المتغيرات ماعدا المتغيرات النصية ولذلك فإن في حالة المتغير النصي نجد أن قيمة المميز لا تكون منشطة تأخذ من خلال المثال رقم ٤ أما باقي المتغيرات فيمكن اعطاء قيمة للتمييز لها، فعلى سبيل المثال متغير النوع Gender يمكننا من تعريف الكود"١" "إلي الذكر والكود"٢" للأنثى وذلك بإتباع الخطوات التالية:

نختار Variable view ⇦ وبعد ذلك عند المتغير الذي نريد أن نعطي له قيمة مميز له وليكن متغير النوع فنقف على الخلية الموجودة في العمود الذي يحمل عنوان "Value" وصف متغير النوع فنجده أن الخلية مكتوب بها فنقوم بنقر على النقاط التي تحمل اللون الرمادي فيظهر مربع الحوار الآتي:



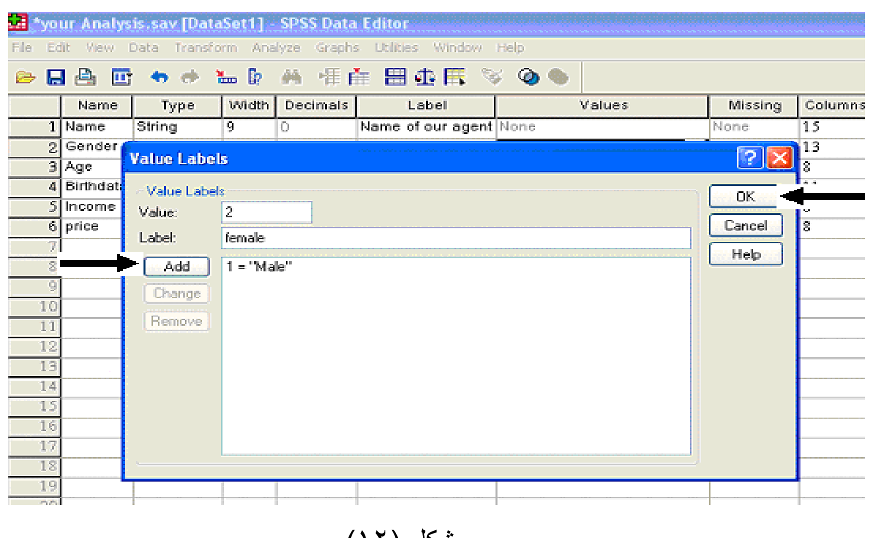
شكل (١٠)

فعندما نرغب في إضافة الكود" ١" للذكر و٢ للأنثى فإننا نكتب أمام الخانة Value القيمة "١" ونكتب أمام الخانة Label ذكراً أو Male فتظهر الشاشة التالية:

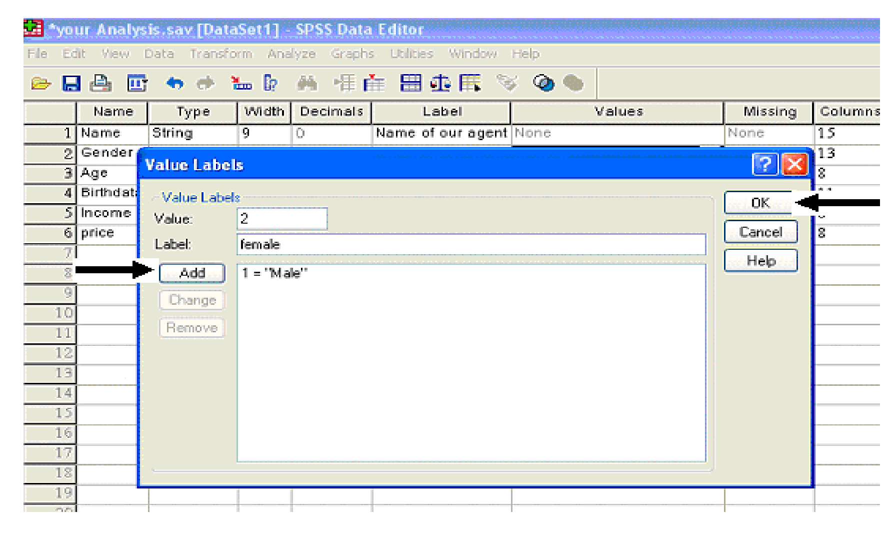


شكل (١١)

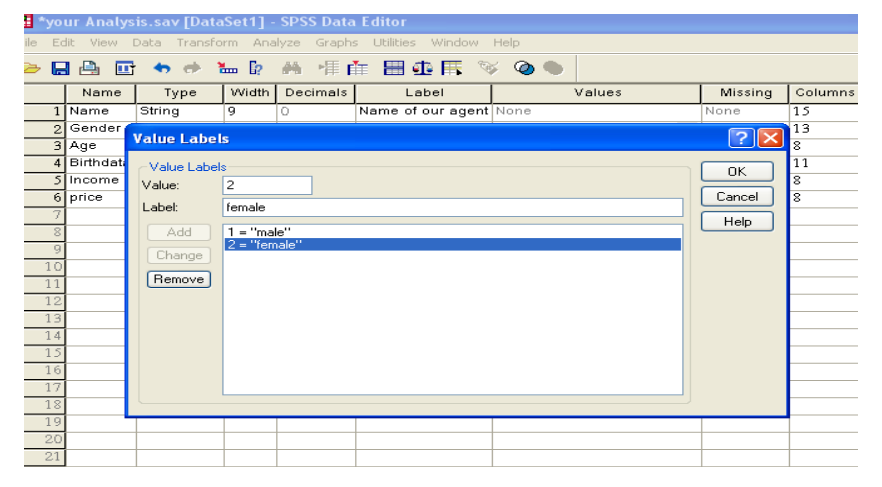
وبعد ذلك ننقر Add وذلك لإضـافة الكود الخـاص بـالأنثى فعندما ننقر Add تظهر الشاشة فنكرر الخطوة السابقة كما فعلنا بالنسبة للكود الذكر ولكننا هنا في خانة Value نضع القيمة '2' وفي خانة Label نضع Female وفي حالة الانتهاء ننقر على Ok ونلاحظ أننا يمكننا حذف أي كود لا نرغب بـه فقط ننقر على الكود المراد حذفه وبعدها النقر على Remove الأشكال الآتية توضح ذلك.



شكل (١٢)



شكل (١٣)



ملاحظات:

- يمكن أن يصل طول عنوان القيمة إلى ٦٠ رمزاً.
- يمكن أن يكون العنوان قيما عددية ليمثل قيما غير عددية مثلا الرقم ١ عنوان للذكور والرقم ٢ عنوان للإناث.
- يمكن إظهار عنوان القيمة view variable أما بالنقر على الأيقونة في شريط الأدوات أو بتأشير Value label من قائمة view في شريط القوائم.

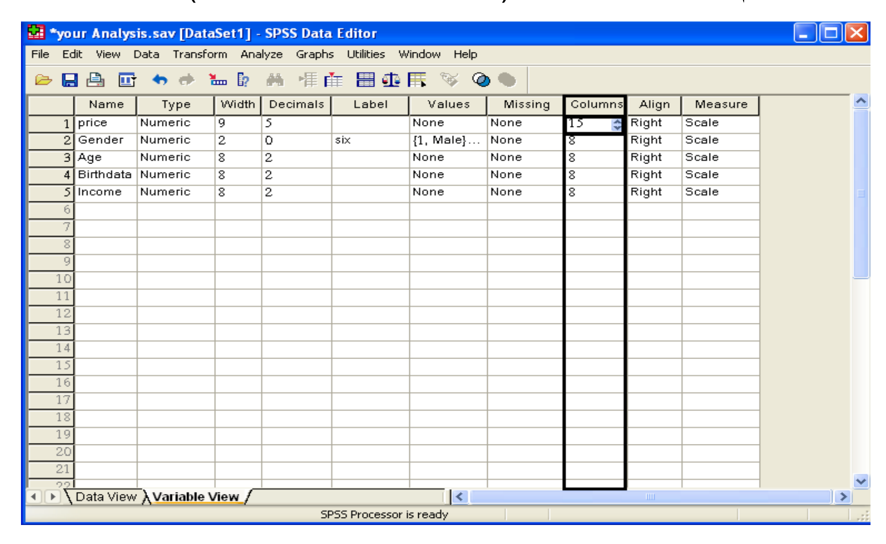
⌖ يظهر عنوان القيمة بدلاً من القيمة نفسها في مخرجات (جدول) برنامج SPSS ويعتبر هذا من أهمية توضيح عنوان للقيم في المتغيرات الاسمية أو الترتيبية.

القيم المفقودة (Missing Values):

القيم المفقودة هي قيم موجودة أصلا، ولكننا لا نرغب إدخالها في التحليل الإحصائي كونها مثلاً قيماً شاذة أو أن نوع السؤال لا ينطبق على المستجيب.

عرض العمود (Column width):

يمكن تحديد عرض العمود لمتغير معين بالوقوف على الخلية الواقعة ضمن العمود المعنون Column في ورقة variable view حيث يمكن زيادة أو تقليل عرض العمود بواسطة الأسهم إلى الأعلى أو الأسفل (أو كتابة عرض العمود مباشرة).



شكل (١٤)

ملاحظات:

⌖ عرض العمود Column يمثل عدد الرموز المخصصة للعمود ويجب أن يكون عرض العمود أكبر أو يساوي عرض المتغير المضمن فيه.

⌖ يمكن تغير عرض العمود لمتغير معين في ورقة Data view مباشرة عن طريق نقر وسحب حدود العمود Clicking and dragging.

محاذاة النص (Alignment):

لضبط محاذاة النص داخل خلايا المتغير انقر الخلية التابعة لمتغير معين في ورقة variable view الواقعة ضمن العمود المعنون Align ثم انقر السهم المتجه للأسفل لاختيار أمر مما يلي:

Left: لمحاذاة النص إلى يسار الخلية.

Center: لمحاذاة النص إلى وسط الخلية.

Right: لمحاذاة النص إلى يمين الخلية.

علماً أن المحاذاة الافتراضية هي (Right).

أنواع وحدات القياس (Measures Types):

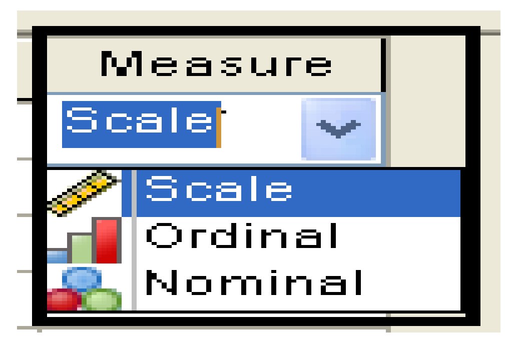
في البداية يجب تعريف أنواع وحدات القياس للبيانات قبل الشروع في تعريف ذلك لبرنامجك الإحصائي.

أنواع وحدات القياس:

- Nominal : يستعمل لقياس المتغيرات الاسمية وهي متغيرات لها عدد من الفئات دون أفضلية لإحداها على الأخرى ولذلك لا يمكن ترتيبها تصاعديا أو تنازليا كما لا يمكن اجراء العمليات الحسابية على هذا النوع من المتغيرات مثل / تقسيم المجتمع إلى ذكور وإناث.
- Ordinal : يستخدم لقياس المتغيرات الترتيبية وحيث أن هذا المتغير ذو عدد محدد من الفئات يمكن ترتيبها تصاعديا أو تنازلياً ولكن لا يمكن تحديد الفروق بينها بدقة مثلا تقدير طالب في الامتحان قد يكون" ممتاز ، جيد جدا ، جيد ، ... الخ" ويمكن أن يكون المتغير رمزيا أو عددياً على أنه يفضل الأخير (عددي).
- Scale : ويستخدم لقياس المتغيرات الكمية في فترة أو نسبة كدخل والعمر و الأسعار ... الخ.

ولغرض تعريف مقياس لمتغير معين انقر خلية المتغير التي تقع ضمن عمود measure في ورقة variable view حيث يظهر ثلاث خيارات scale، ordinal ،

and nominal ونختار نوع المتغير على حسب التعاريف السابقة ومن ثم تظهر شاشة variable view كما يلي:

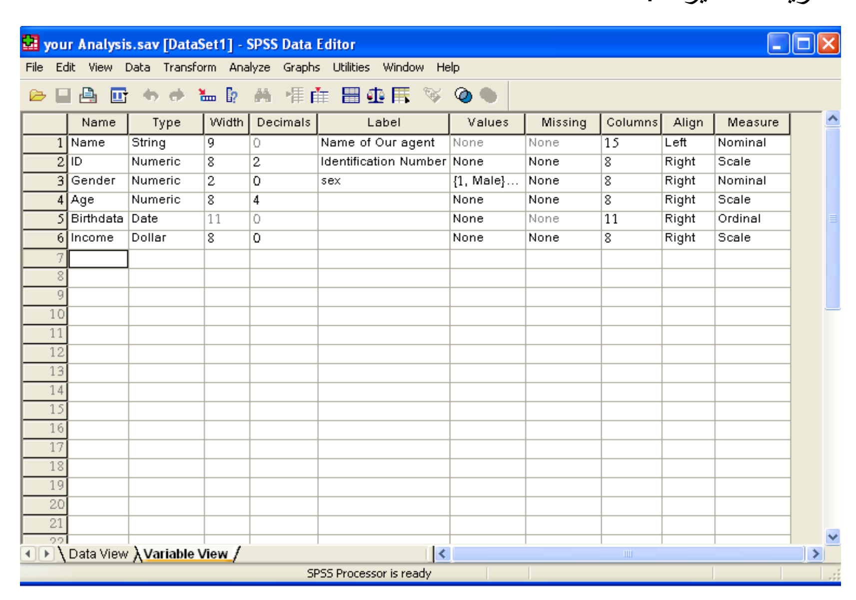


شكل (١٥)

وفي النهاية تظهر شاشة variable view.

نريد الآن ادخال بيانات الجدول (١) الموجودة في مثال (١) في بداية هذا الفصل.

بعد تعريف المتغيرات:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Name	String	9	0	Name of Our agent	None	None	15	Left	Nominal
2	ID	Numeric	8	2	Identification Number	None	None	8	Right	Scale
3	Gender	Numeric	2	0	sex	{1, Male}...	None	8	Right	Nominal
4	Age	Numeric	8	4		None	None	8	Right	Scale
5	Birthdata	Date	11	0		None	None	11	Right	Ordinal
6	Income	Dollar	8	0		None	None	8	Right	Scale

شكل (١٦)

إدخال البيانات (Entering Data):

الآن وصلنا إلى مرحلة إدخال البيانات استعداداً لإجراء التحليل الإحصائي.

يتم إدخال البيانات كما يلي :

⌖ نختار الخلية المراد إدخال قيمة المتغير فيها.

⌖ نكتب القيمة المراد إدخالها في الخلية في مكان اسمه محرر الخلية Cell Editor موجود تحت شريط الأدوات.

⌖ عند الانتقال للخلية التالية بالأسهم أو النقر على الأمر Enter تكتب القيمة المدخلة تلقائياً.

⌖ عند حدوث أية أخطاء في عملية الإدخال يمكن إجراء التصحيح كالآتي :

⌖ ننتقل إلى الخلية المراد تصحيحها ويتم النقر عليها.

⌖ نكتب القيمة الصحيحة في محرر الخلية.

ملاحظة : ممنوع إدخال تعبيرات حسابية (جمع / ضرب / ...).

والآتي شكل data view بعد إدخال بيانات جدول رقم (١) فيه:

*your Analysis.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

7 : Income — Visible: 6 of 6 Vari

	Name	ID	Gender	Age	Birthdata	Income	var	var	var	var	var
1	Mohammed	1.00	1	38.0000	01.01.1970	$1,000					
2	Mervat	2.00	2	27.0000	20.08.1980	$500					
3	Ahmad	3.00	1	48.0000	15.03.1960	$2,000					
4	Heba	4.00	2	22.0000	15.04.1985	$300					
5	Nabil	5.00	1	18.0000	01.01.1990	$200					
6	Noha	6.00	2	58.0000	20.07.1949	$3,000					

Data View / Variable View

SPSS Processor is ready

شكل (١٧)

عمليات مختلفة على البيانات:

- **تحديد المتغيرات والحالات:**

⌖ لاختيار (تحديد أو تظليل) متغير ما variable انقر الخلية التي تحتوي على اسم المتغير في أعلى العمود في ورقة Data view بزر الماوس الأيسر.

⌖ لاختيار حالة case بأكلمها انقر الخلية الحاوية على رقم الحالة في ورقة Data view بزر الماوس الأيسر.

- لاختيار مجموعة من المتغيرات المتجاورة: انقر الخلية الحاوية اسم المتغير الأول.
- النقر على مفتاح shift.
- انقر الخلية الحاوية على اسم المتغير الأخير مع استمرار النقر على مفتاح shift.
- لاختيار مجموعة من المتغيرات المتباعدة انقر بزر الماوس الأيسر الخلية الحاوية اسم المتغير الأول لاختياره.
- انقر على مفتاح Ctrl مع استمرار النقر عليه (انقر الخلية الحاوية اسم المتغير الثاني لاختيارها وهكذا بالنسبة لباقي المتغيرات التالية).
- بنفس الطريقة المستخدمة لاختيار مجموعة من المتغيرات المتجاورة أو غير متجاورة يمكن استعمالها لاختيار مجموعة من الحالات Cases المتجاورة أو غير المتجاورة.

- **إدخال (حشر) حالة جديدة** (Insert a New Case):

نستطيع حشر حاله بين حالتين موجودتين كالتالي:

- ننتقل إلى (الصف) أسفل المكان المراد إضافة الحالة فيه ويتم النقر.
- من القائمة Data ننقر على الأمر Insert Case فتظهرحالة جديدة خالية.
- ندخل البيانات المراد إدخالها إلى الخلايا عن طريق محرر الخلية.

- **إدخال (حشر) متغير جديد** (Insert new variable):

الطريقة الأولى:

من شريط القوائم اختر القائمة Data ثم نختار Insert Variable فيضاف متغير جديد إلى يسار المتغير الحالي حيث يمكن تغيير الاسم .

الطريقة الثانية:

الوقوف على المتغير المراد إضافته على يسار المتغير وثم النقر بزر الماوس الأيمن واختيار insert variable وتسمية المتغير بعد ذلك في شاشة variable view بنفس الطريقة السابقة.

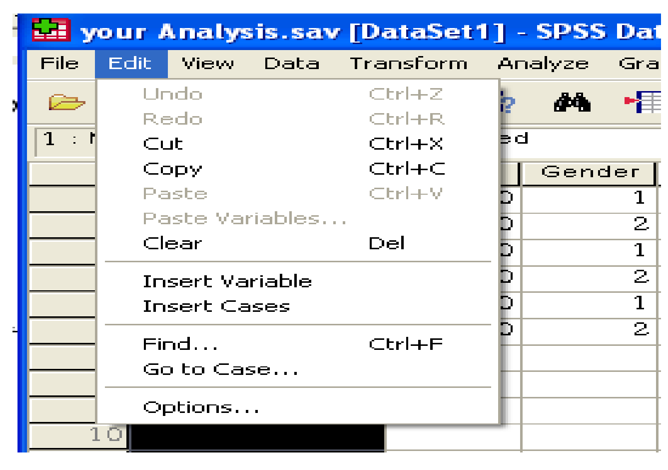
▪ **حذف صف** (Delete Row):

انقر اسم الحالة في ورقة Data View بزر الماوس الأيسر.

من القائمة Edit انقر على Clear تتم الإزالة الفورية للحالة. ويمكن استخدام الأمر Delete بعد اختيار المتغير مباشرة .

▪ **حذف عمود** (Delete Coulmn):

يتم حذف المتغير كالآتي حدد المتغير المراد حذفه من القائمة Edit التي تتضمن القائمة الآتية:



شكل (١٨)

انقر على Clear ويمكن استخدام الأمر Delete .

▪ **حذف صف** (Delete Case):

بنفس طريقة حذف المتغيرات يمكننا حذف الحالات الغير مرغوب فيها كالأتي:

انقر اسم الحالة في ورقة Data View بزر الماوس الأيسر.

من القائمة Edit انقر على Clear تتم الإزالة الفورية للحالة . ويمكن استخدام الأمر Delete بعد اختيار المتغير مباشرة .

▪ **نسخ وقص ولصق الحالات والمتغيرات (copy ، cut ، Paste of cases and variable)**

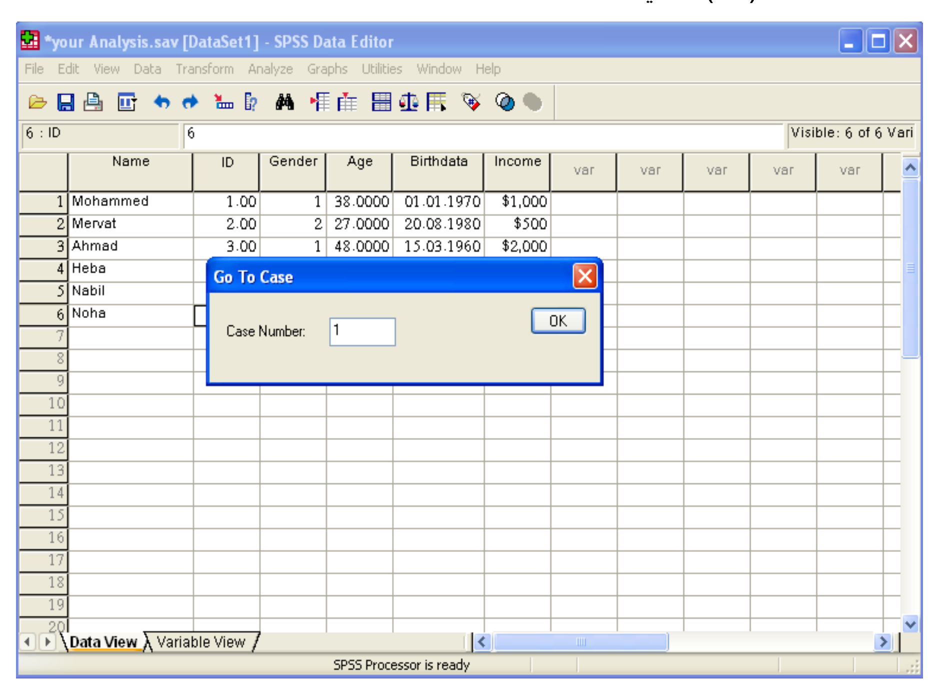
لعمل نسخة من متغير معين Copy نتبع الخطوات التالية:

⌖ حدد اسم المتغير ثم من شريط القوائم نختار edit ومن ثم اختيار copy من القائمة المدرجة انظر شكل (١٩) ويمكن النقر مباشرة على ctrl+c لإجراء عملية النسخة بشكل سريع نحدد بعدها المتغير المراد النسخ فيه.

⌖ من شريط القوائم نختار edit ومن ثم اختيار Past ويمكن النقر مباشرة على ctrl+v لإجراء عملية اللصق بشكل سريع. وعملية القص واللصق تتم بنفس الطريقة ، كما أن عملية نسخ ولصق وقص الحالات أيضا تتم بنفس الطريقة مع تبديل اختيار العمود (المتغير) بالصف (الحالة).

▪ **البحث عن حالات** (Go To Case):

⌖ انقر الأمر Go To Case من القائمة Data فيظهر المربع الحواري (انظر الشكل (١٩) التالي:



الشكل (١٩)

⌖ في المربع Case Number اكتب رقم الحالة التي تريد الانتقال إليها.

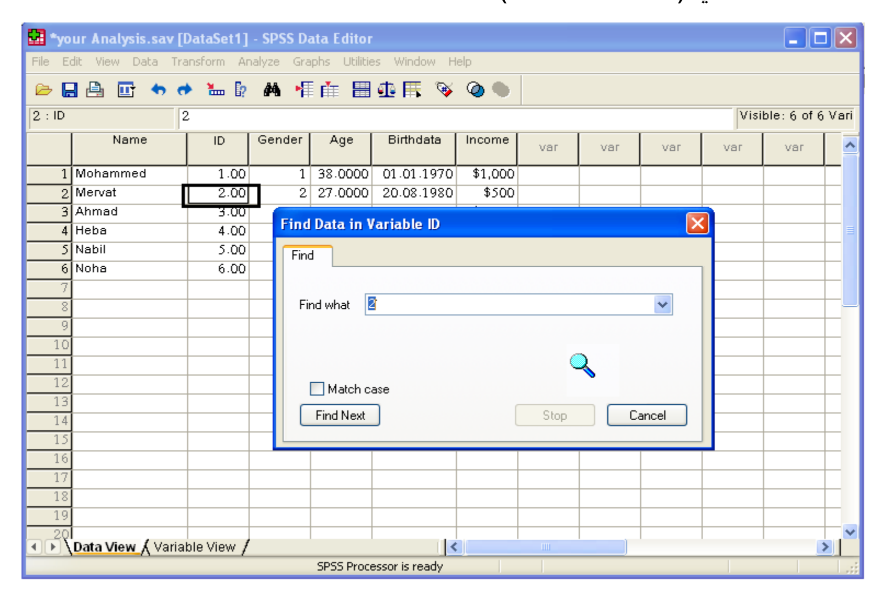
⌖ انقر Ok فيتم الانتقال إلى الحالة المحددة.

▪ **البحث عن قيمة** (Finding Values Finding Values):

تتم كالآتي:

⌖ ننتقل إلى قمة العمود الخاص بالمتغير.

⌖ انقر الأمر Find من القائمة Edit فيظهر المربع الحواري Search for Data الآتي (أنظر الشكل ٢٠):



الشكل (٢٠)

⌖ في المربع الحواري حدد الرقم المطلوب البحث عنه في المربع Find What ونلاحظ أننا إذا أردنا البحث عن القيمة" 2 " فإننا نكتب القيمة" 2 " في خانة Find what وبعدها النقر على Find next فنلاحظ أن البرنامج يقف على أول حالة أو متغير لدية القيمة" 2 "ويحددها كما في الشكل السابق.

الفصل الثانى

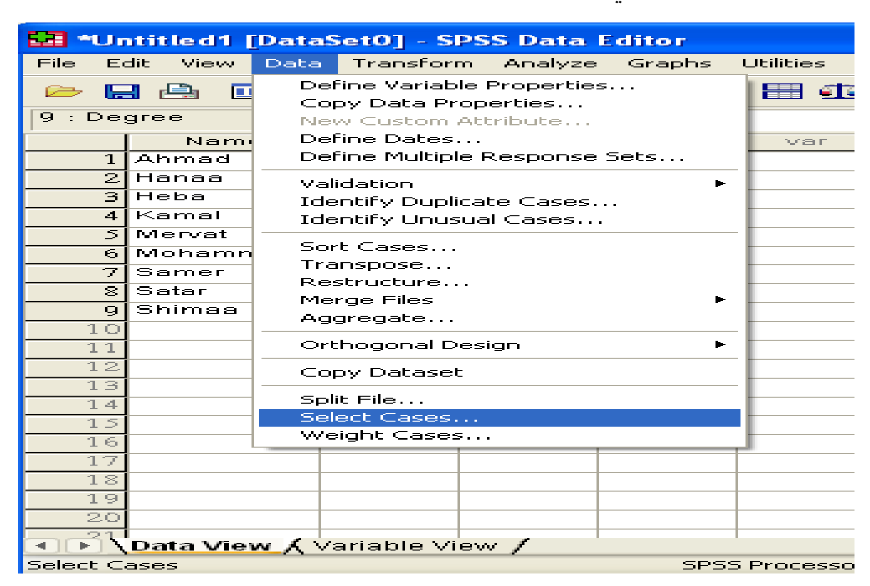
معالجــة البيانات

Data Processing

استعرضنا في الفصل السابق طريقة تعريف المتغيرات وإدخال البيانات محل الدراسة لبرنامج SPSS تمهيداً لإجراء التحليل الإحصائي عليها. في هذا الفصل نناقش بعض طريق معالجة البيانات وإجراء التحليل الإحصائي عليها.

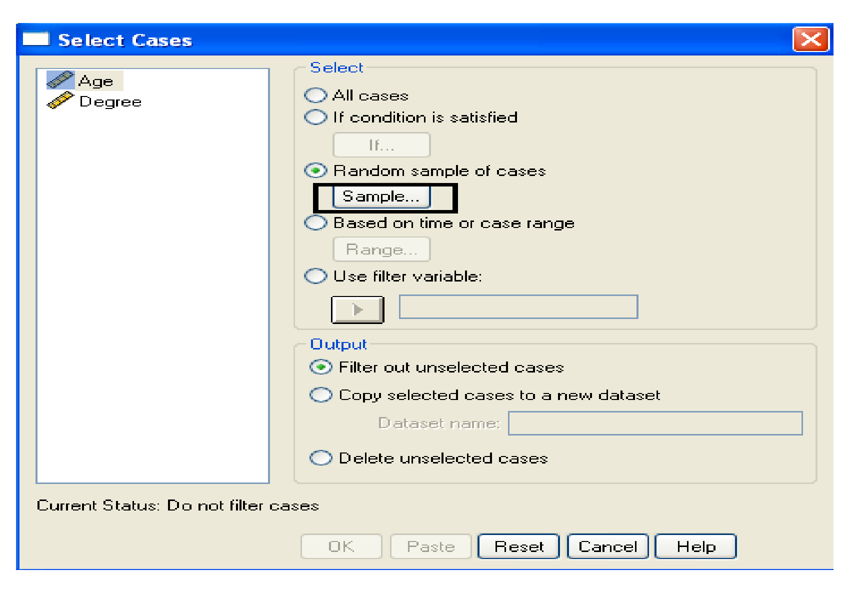
اختزال (تخفيض) حجم البيانات Reducing sample size :

نفرض أننا لدينا ٢ مليون بيان تم الحصول عليها من المشاهدات و لتخفيض حجم هذة البيانات، نقوم بإختيار عينة عشوائية منها لإجراء التحليل الإحصائي عليها. نفرض أنه تم اختيار ١٠٠ الف مشاهدة بطريقة العينة العشوائية البسيطة ولضمان عدم وجود تحيز في الاختيار نقوم باختيار Select Cases من القائمة Data كما بالشكل التالي:



شكل (١)

بعدها تظهر الشاشة التالية :



شكل (٢)

فنقوم باختيار Random sample of cases والنقر بعدها على مربع الحوار Sample و لأننا نختار ١٠٠ الف من ٢ مليون بيان فإننا نحدد حجم العينة بتقريب بنسبة ٥% من البيانات ومن ثم تظهر الشاشة التالية:

شكل (٣)

ترتيب البيانات Sort Data :

نقصد بعملية ترتيب البيانات بأن نرتب بيانات أحد المتغيرات (أو أكثر من متغير) في الملف ثم نرتب بقية المتغيرات تبعا لذلك. لترتيب بيانات ملف نذكر بعض الملاحظات الهامة :

- أن ترتيب بيانات الملف تتم بالنسبة لمتغير أو أكثر وليست مطلقة.
- أن عملية الترتيب تعني نقل الحالات الموجودة في الملف إلى أعلى أو إلى أسفل .
- لا يجوز نقل قيمة معينة (أو قيم) من حالة من مكانها إلى مكان آخر بل يجب نقل الحالة كلها.

مثال (١):

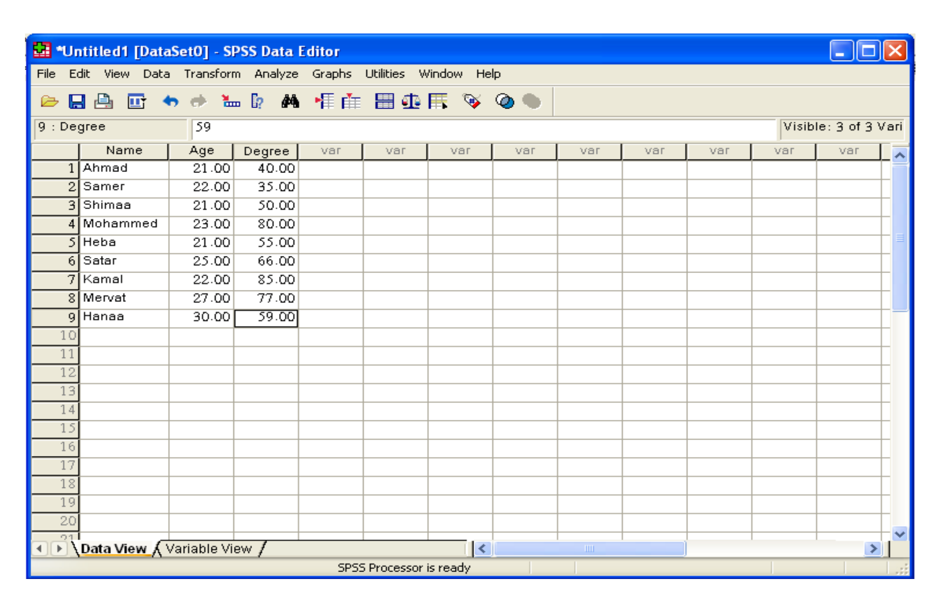
لدينا بيانات الملف التالي لخرجين أحد كليات العلوم:

جدول (١)

Name	Age	Total degree
Ahmad	21	40
Samer	22	35
Shimaa	21	50
Ali	23	80
Heba	21	55
Satar	25	66
Kamal	22	85
Mervat	27	77
Hanaa	30	59

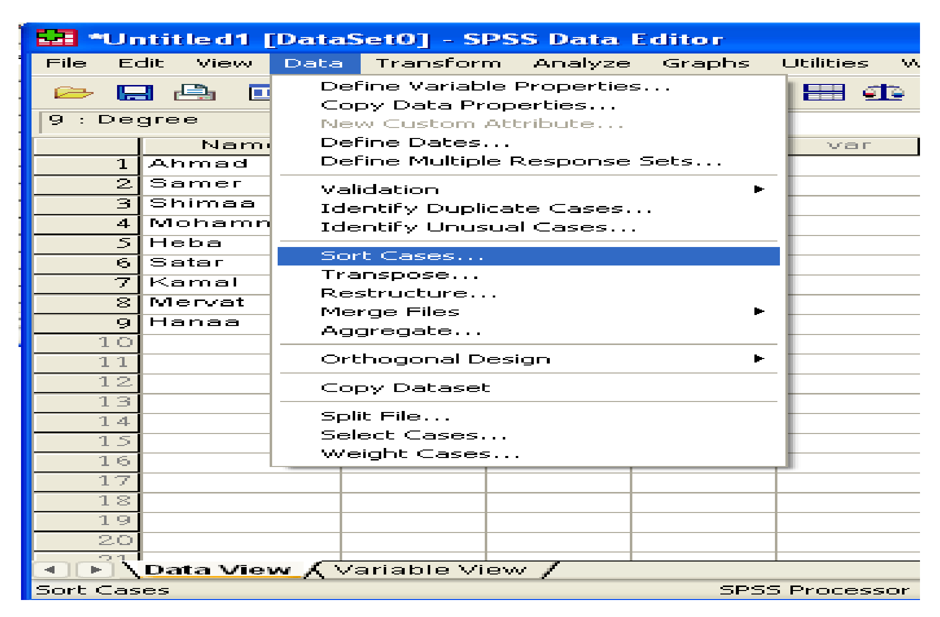
لتنفيذ ترتيب البيانات باستخدام البرنامج نتبع الخطوات التالية :

نقوم بإدخال البيانات في البرنامج فتظهر الشاشة التالية:



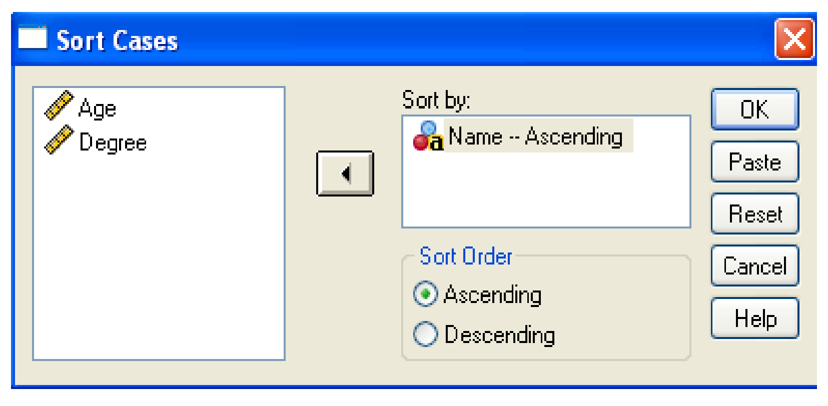
شكل (٤)

انقر Data فتظهر الشاشة التالية:



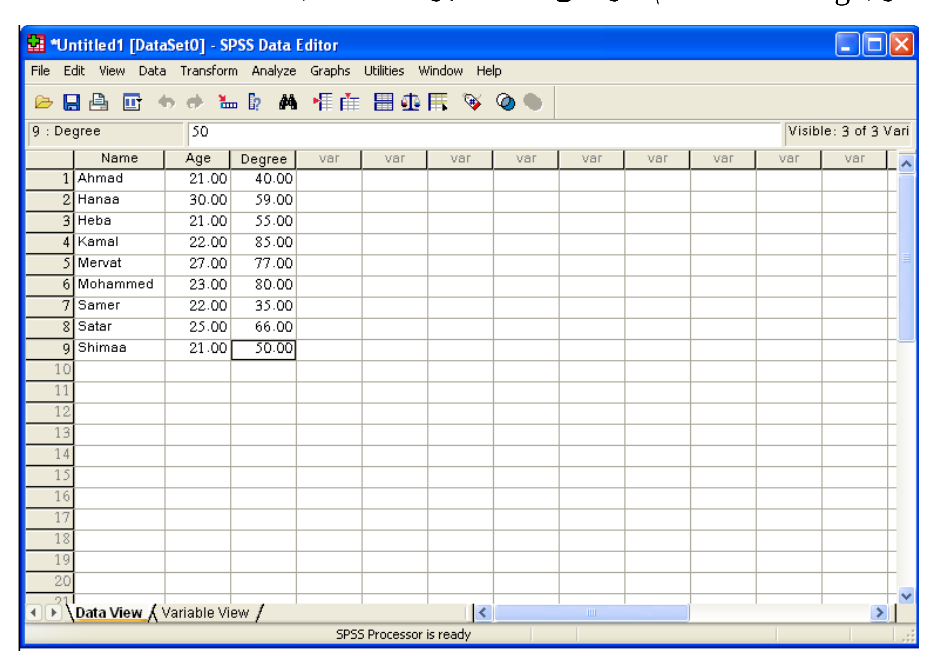
شكل (٥)

ثم نقوم باختيار Sort Cases، يفتح الصندوق الحواري Sort Cases وبه أسماء جميع المتغيرات. نختار المتغير الذي سيتم على أساسه الترتيب وهو Name وينقل إلى المربع Sort By



الشكل(٦)

نختار نوع الترتيب من Sort Order هل الترتيب سيكون تصاعديا Ascending أو تنازليا Descending ثم ننقر على OK فتظهر الشاشة التالية:



شكل (٧)

نلاحظ بعد تنفيذ الأمر السابق أن البرنامج سيقوم بترتيب البيانات تصاعديا.

دمج الملفات Marge Files :

قد تستدعي الحاجة إجراء عملية ضم ملفات بيانات SPSS و في الحقيقة البرنامج يسمح بإجراء عملية الدمج بين الملفات باستخدام حالات إضافية في نهاية ملفك أو دمج متغيرات إضافية لكل الحالات الموجودة في ملفك:

- الدمج بإضافة حالات جديدة Adding Cases :

هذة العملية تسمح بدمج الملفات التي تحتوى علي نفس المتغيرات ولكن حالات مختلفة:

مثال (٢) :

نفرض لدينا الملف الأول وبه درجات طالبين لأربع مواد Your data since 1990، وتم حفظ البيانات التاليه في مجلد باسم Our Book:

جدول (٢)

Name	Math	Chem	Physc	Music
Ahmad	100	90	95	87
Hanaa	95	87	90	85

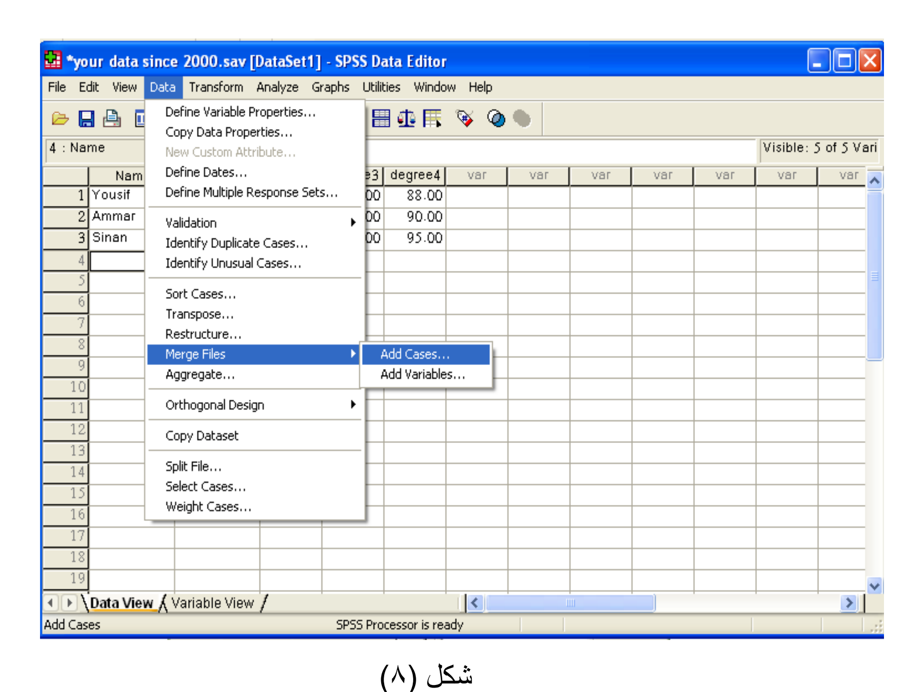
ولدينا الملف الثاني وبه درجات ٣ طلاب لأربع مواد Your data since 2000 :

جدول (٣)

Name	Math	Chem	Physc	Paint
Yousif	85	90	77	88
Ammar	95	83	82	90
Sinan	90	92	86	95

يمكننا من دمج الملفين معاً باستخدام الخطوات التالية:

افتح أحد الملفين في شاشة المحرر وليكن Your data since 2000 ومن قائمة Data اختر الأمر Merge File فيظهر الشكل التالي:



شكل (٨)

نقوم باختيار add cases بنقر عليها ومن ثم تظهر الشاشة التالية:

Add Cases to your data since 2000.sav [DataSet1]

Select a dataset from the list of open datasets or from a file to merge with the active dataset

An open dataset

An external SPSS data file

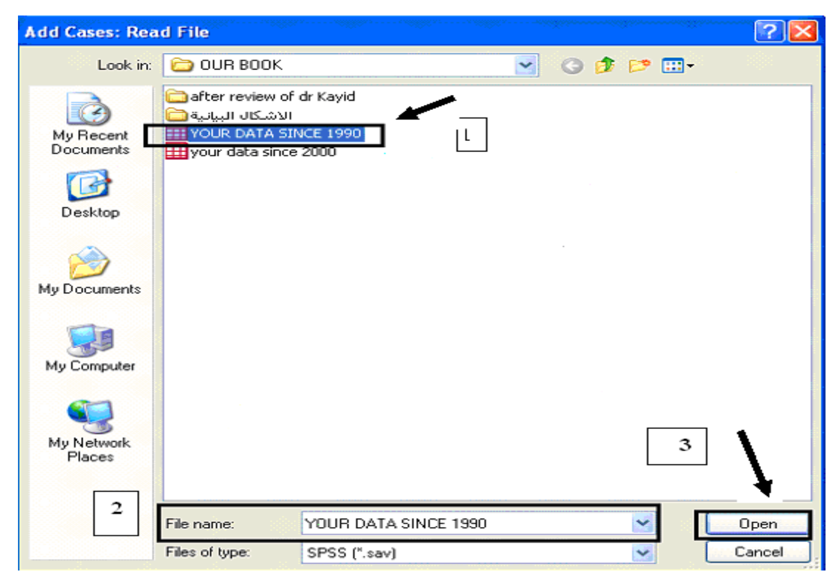
Browse...

Non-SPSS data files must be opened in SPSS before they can be used as part of a merge.

Continue Cancel Help

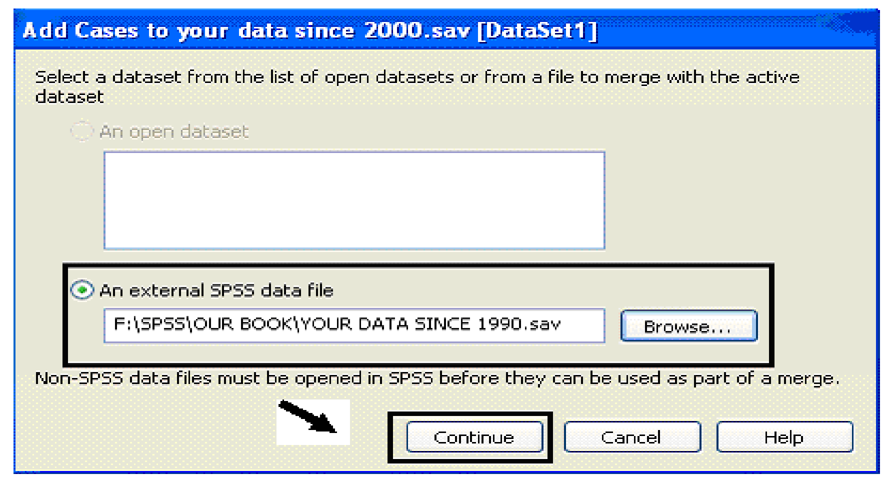
شكل (٩)

ولأننا ندمج ملف آخر من ملفات SPSS نقوم بتحديد An external SPSS data file وبعدها نقوم بنقر Browse لتحديد مكان الملف الآخر ومن ثم تظهر الشاشة التالية:



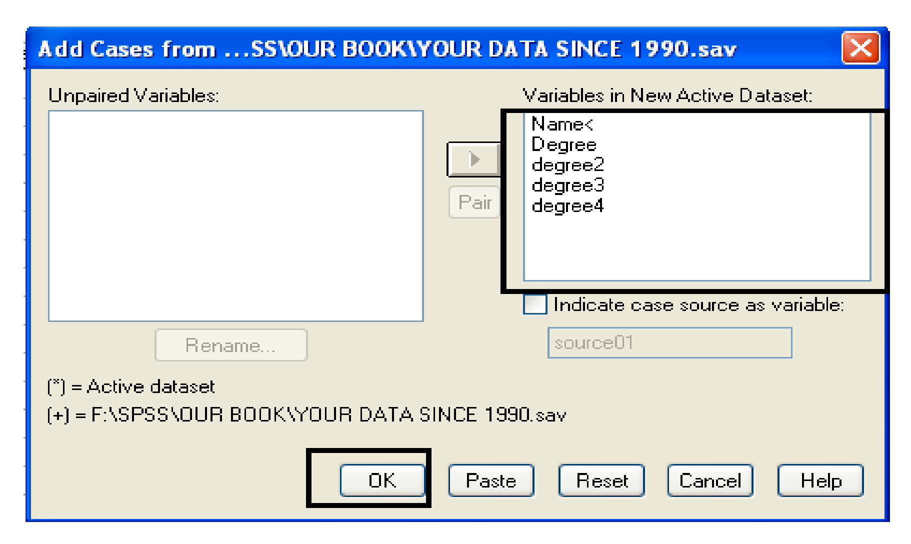
شكل (١٠)

وبعدها تظهر الشاشة التالية موضح بها مكان الملف المراد دمجه مع الملف الآخر:



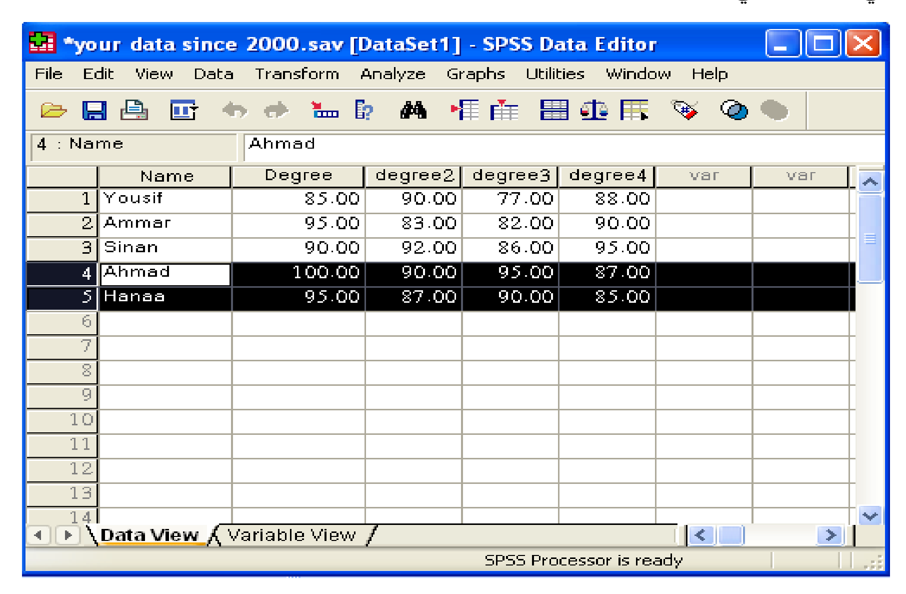
شكل (١١)

وبنقر continue تظهر الشاشة التالية موضح بها متغيرات ملف your data since 1990 وفي أسفل الشاشة ننقر Ok:



شكل (١٢)

ومن ثم يتم دمج ملف your data since 1990 مع الملف your data since 2000 ونلاحظ أن بيانات الملف المدمج تضاف في نهاية ملف your data since 2000 كما في الشكل آلاتي:



	Name	Degree	degree2	degree3	degree4	var	var
1	Yousif	85.00	90.00	77.00	88.00		
2	Ammar	95.00	83.00	82.00	90.00		
3	Sinan	90.00	92.00	86.00	95.00		
4	Ahmad	100.00	90.00	95.00	87.00		
5	Hanaa	95.00	87.00	90.00	85.00		
6							
7							
8							
9							
10							
11							
12							
13							
14							

شكل (١٣)

▪ **الدمج بإضافة متغيرات جديدة** Adding Variables :

هذة العمليـة تسـمح بـدمج الملفـات التـي تحتـوي علـى نفس الحـالات ولكـن متغيـرات مختلفة:

مثال (٣) :

نفرض لدينا الملف الأول وبه بيانات أولية عن عملاء شركة النصر للسيارات:

جدول (٤)

Name	ID	AGE	Gender
Mervat Mahdy	123	28	1
Ahmad Ali	122	40	2
Mona Mohammed	145	33	1

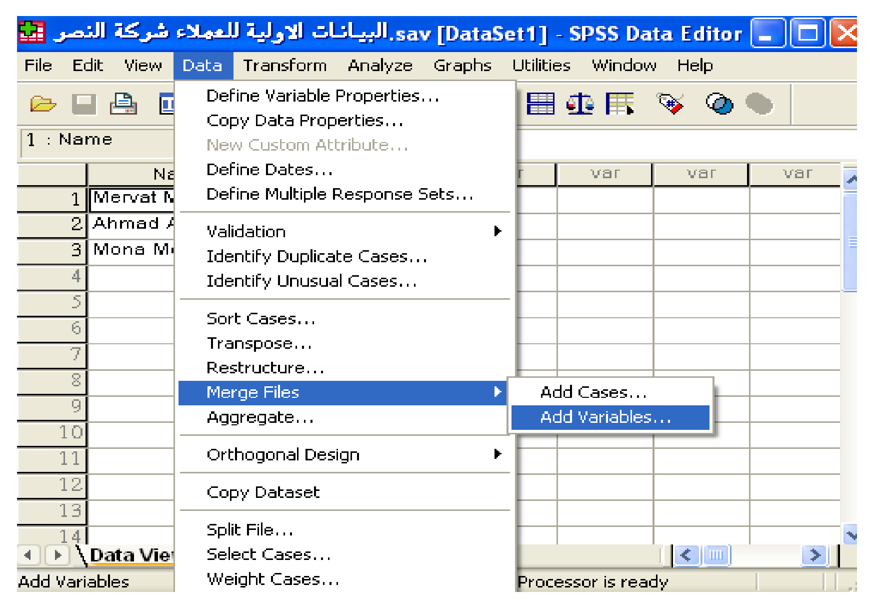
و لدينا الملف الثاني وبه بيانات تفصيلية عن نفس العملاء لنفس الشركة:

جدول (٥)

Name	Adress	Income
Mervat Mahdy	Cairo	1000
Ahmad Ali	Alex	2000
Mona Mohammed	Giza	1500

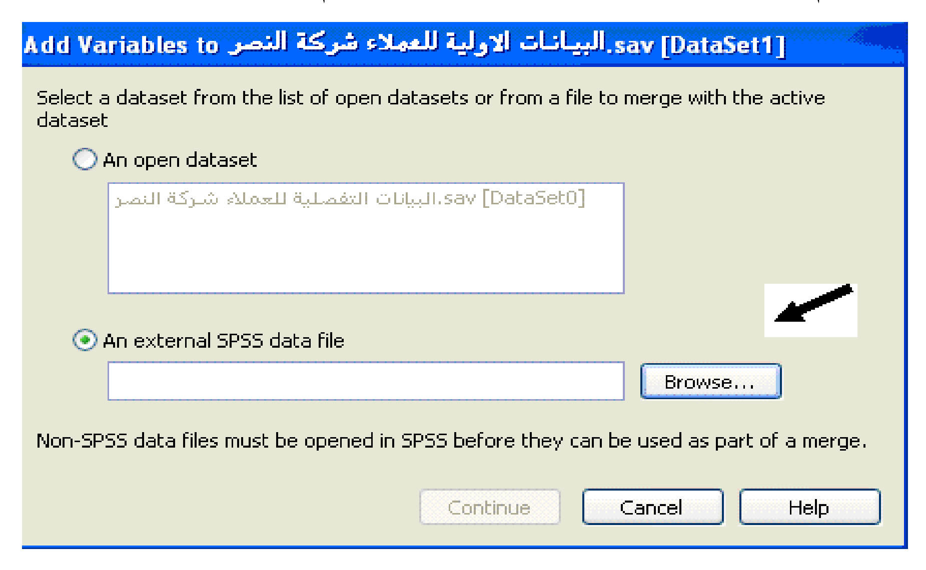
يمكننا من دمج الملفين معاً باستخدام الخطوات التالية:

✓ افتح أحد الملفين في شاشة المحرر وليكن البيانات الأولية لعملاء شركة النصـر ومـن قائمة Data اختر الأمر Merge File فيظهر الشكل التالي:



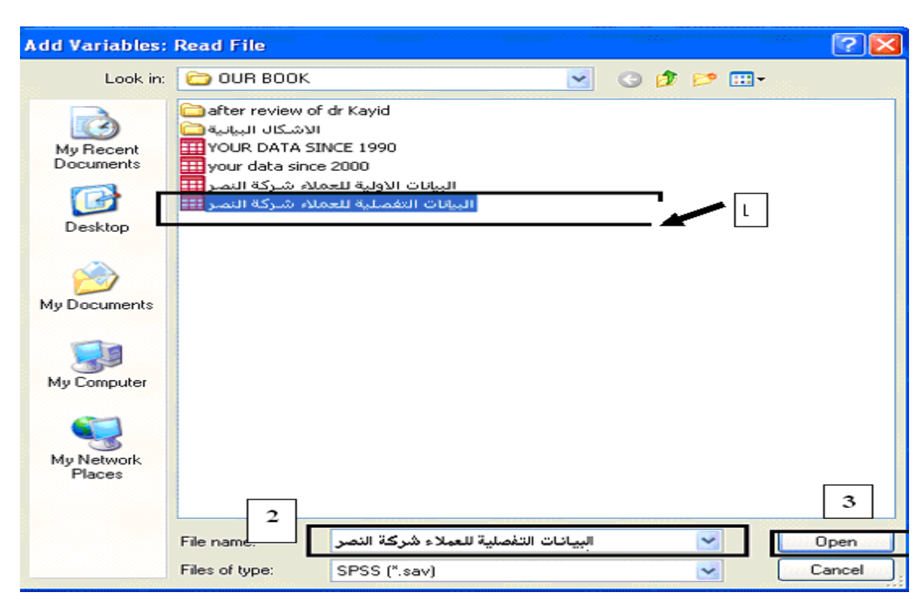
شكل (١٤)

✓ نقوم باختيار add Variables... بنقر عليها ومن ثم تظهر الشاشة التالية:



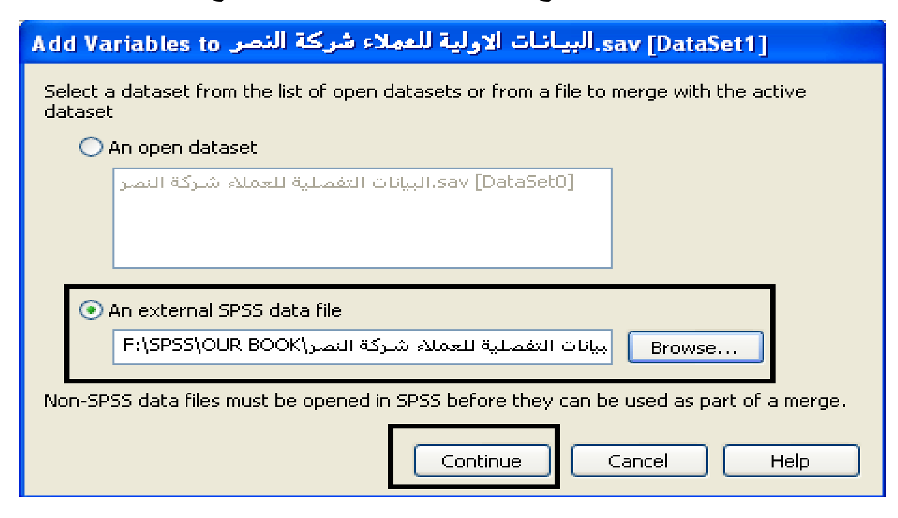
شكل (١٥)

ولأننا ندمج ملف أخر من ملفات SPSS نقوم بتحديد An external SPSS data file وبعدها نقوم بنقر على Browse لتحديد مكان الملف الآخر ومن ثم تظهر الشاشة التالية:



شكل (١٦)

وبعدها تظهر الشاشة التالية موضح بها مكان الملف المراد دمجه مع الملف الآخر:

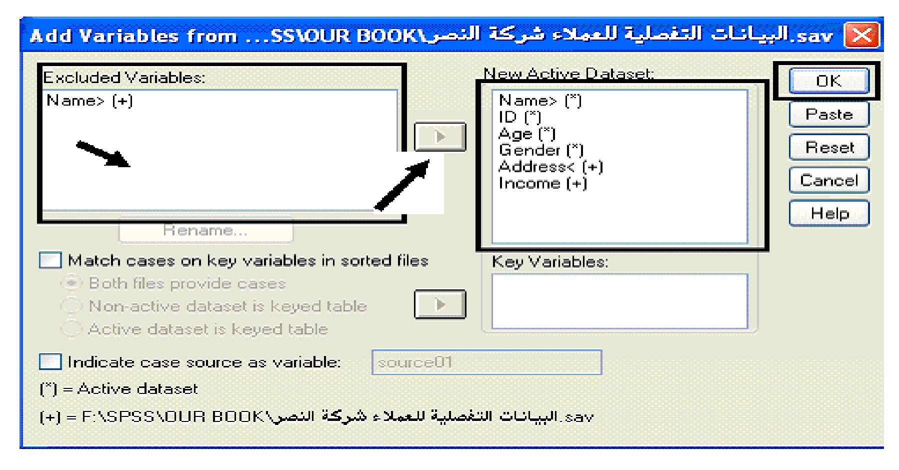


شكل (١٧)

ومع نقر continue تظهر الشاشة التالية وتتضمن ما يلي:

New Active Dataset: وهي قائمة بها جميع المتغيرات في كلا من الملفين التي لا تتشابه من ناحية الاسم في كلا من الملفين.

Excluded Variables: وهي قائمة بأسماء المتغيرات التي تستبعد من الملف المدمج الجديد وعادة ما تتضمن أسماء المتغيرات المشابة أسمائها وبعد ذلك النقر على Ok:



شكل (١٨)

ومن ثم يتم دمج ملف البيانات الأولية للعملاء شركة النصر مع الملف البيانات التفصيلية للعملاء شركة النصر ونلاحظ أن بيانات الملف البيانات الأولية للعملاء شركة النصر تضاف إليه المتغيرات الجديدة إلى يمين المتغيرات القديمة طالما هو الملف المدمج فيه كما في الشكل الآتي:

*البيانات الاولية للعملاء شركة النصر.sav [DataSet1] - SPSS Data Ed...

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : Address Cario

	Name	ID	Age	Gender	Address	Income	var
1	Mervat Mahdy	123.00	28.00	1	Cario	1000	
2	Ahmad Ali	122.00	40.00	2	Alex	2000	
3	Mona Mohammed	145.00	33.00	1	Giza	1500	
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							

Data View | Variable View

SPSS Processor is ready

شكل (١٩)

تحوير ملفات Transpose Files :

نعني بتحوير الملف جعل الصفوف أعمدة والأعمدة صفوف أو بمعنى آخر تحويل الحالات لمتغيرات والمتغيرات لحالات، بشرط أن تكون كلها من نفس النوع.

جدول (٦)

Name	ID	Age	Gender	Address	Income
Mervat Mahdy	123	28	1	Cairo	1000
Ahmad Ali	122	40	2	Alex	2000
Mona Mohammed	145	33	1	Giza	1500

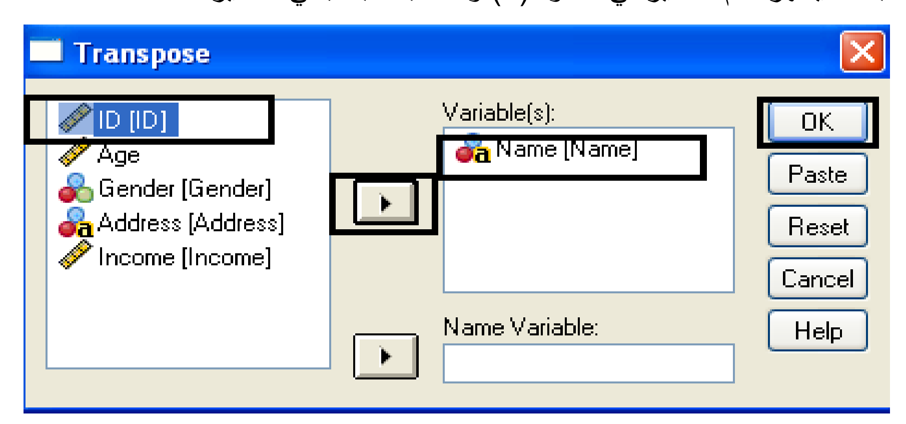
مثال (٤) :

إذا أردنا تحوير هذا الملف سيكون لدينا ثلاثة متغيرات وحالتين في الملف الجديد.

يمكن إجراء عملية التحوير باستخدام SPSS كالآتي :

- نفتح الملف الذي به البيانات فتكون البيانات في الشكل (١٩) السابق.
- من القائمة Data ننقر على Transpose فيظهر الصندوق الحواري وبه أسماء المتغيرات الموجودة في الملف.
- ننقل المتغيرات إلى مربع Variables بخطوات الآتية:

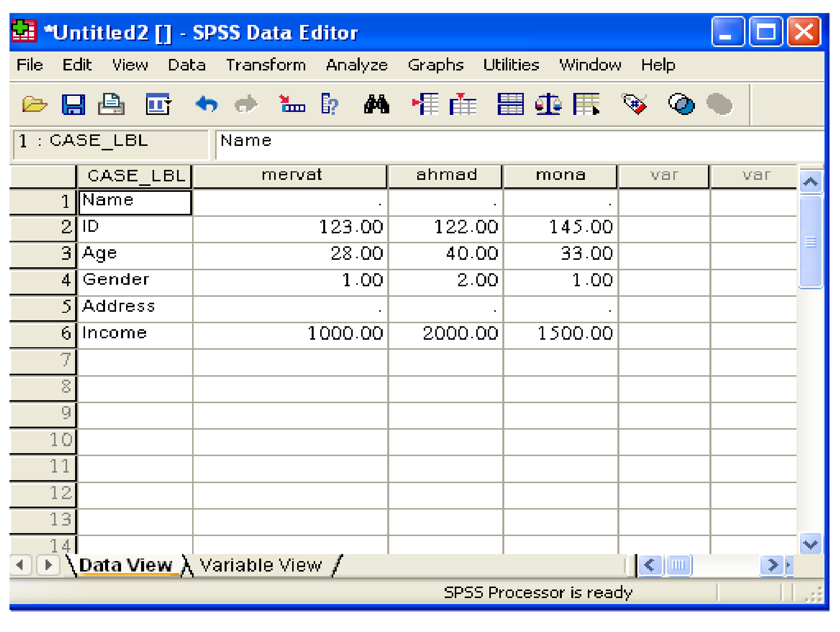
في البداية تحديد المتغير المراد نقله وهو في العمود ١ ثم النقر على مفتاح النقل (٢) بعدها يظهر اسم المتغير في العمود (٣) وهكذا بالنسبة لباقي المتغيرات.



شكل (٢٠)

✓ ننقر Ok لنحصل على الملف المحور.

فيصبح شكل ملف البيانات في الشكل الآتي:



	CASE_LBL	mervat	ahmad	mona	var	var
1	Name	.	.	.		
2	ID	123.00	122.00	145.00		
3	Age	28.00	40.00	33.00		
4	Gender	1.00	2.00	1.00		
5	Address	.	.	.		
6	Income	1000.00	2000.00	1500.00		
7						
8						
9						
10						
11						
12						
13						
14						

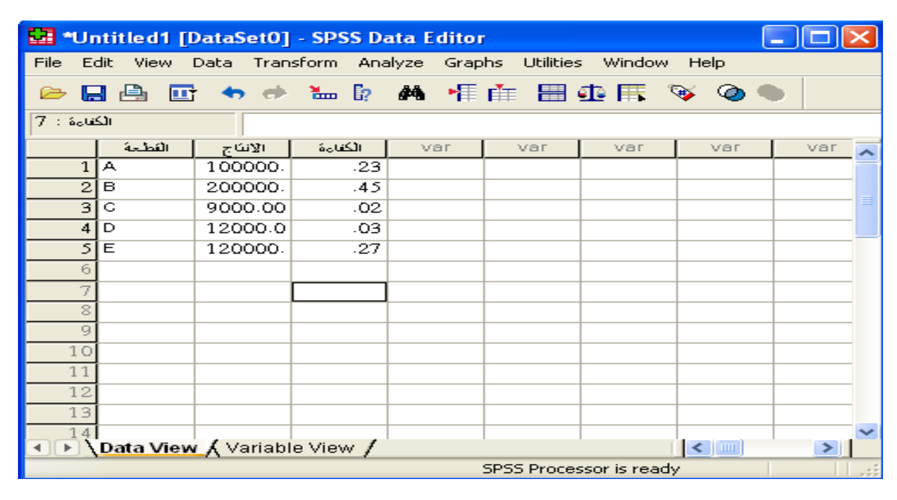
شكل (٢١)

تلخيص الحالات Aggregate :

يستعمل هذا الأمر لتلخيص المعلومات المتعلقة بمجموعة من الحالات cases في حالة تجميعية واحدة وتكوين ملف تجميعي جديد ويمكن أن نجري تلخيص للحالات باستخدام البرنامج كما في المثال التالي:

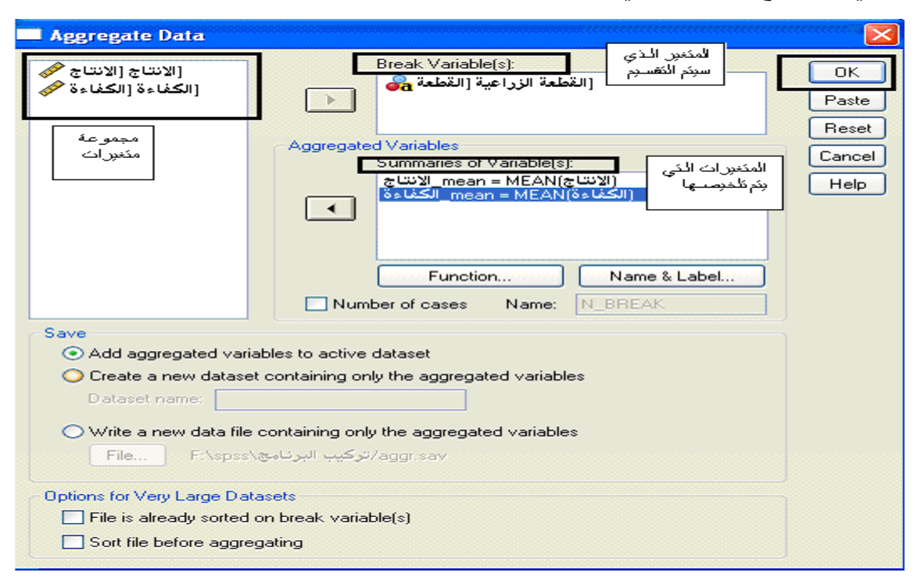
مثال (٥):

الملف الآتي يوضح بيان بإنتاجية ٥ قطع زراعية معينة تنتج محصول القمح ونسبة كفاءة كل قطعة زراعية كما يلي:



شكل (٢٢)

من القائمة Data انقر الأمر Aggregate يظهر الصندوق Aggregate Data التالي الموضح بالشكل التالي:



الشكل (٢٣)

⌖ من قائمة المتغيرات على اليسار اختر المتغير الذي سيتم على أساسه التقسيم وانقله إلى Break Variable ويتم النقل عن طريق زر الانتقال الموجود بجوار Break

Variable من ناحية اليسار الظاهر بشكل [▶] ، ونلاحظ أنه عندما ننقل هذا المتغير فإنه يحذف من عمود مجموعة المتغيرات.

- قم بنقل المتغير (أو المتغيرات) المطلوب اجراء العمليات الإحصائية لها بناء على متغير التقسيم إلى Summaries of variable(s). لاحظ أننا يمكننا الرجوع عن الاختيار باستخدام الزر الآتي الموجود بجوار Summaries of variable(s) من ناحية اليسار الظاهر بشكل [◀].
- انقر على الأمر Function فيظهر لك الصندوق Aggregate Function: Aggregate Data وهو يحتوي على عمليات إحصائية كثيرة ولاحظ انه لا يحق للمستخدم أن يختار أكثر من عملية إحصائية في نفس الوقت (انظر الشكل ٢٤).

Aggregate Data: Aggregate Function

Summary Statistics	Specific Values	Number of cases	
◉ Mean	◯ First	◯ Weighted	Continue
◯ Median	◯ Last	◯ Weighted missing	Cancel
◯ Sum	◯ Minimum	◯ Unweighted	Help
◯ Standard deviation	◯ Maximum	◯ Unweighted missing	

Percentages
- ◯ Above
- ◯ Below

Value:

- ◯ Inside
- ◯ Outside

Low: High:

Fractions
- ◯ Above
- ◯ Below

Value:

- ◯ Inside
- ◯ Outside

Low: High:

الشكل (٢٤)

- انقر على الأمر Continue للعودة للصندوق الأصلي .
- انقر على الأمر Ok للتنفيذ.

لاحظ انه تم إيجاد الوسط الحسابي كما طلبنا في الأمر Aggregate Function.

فصل الملفات Split Files :

يستعمل هذا الأمر لغرض تجزئة (فصل) ملف البيانات لأغراض التحليل الإحصائي .

مثال (٦) :

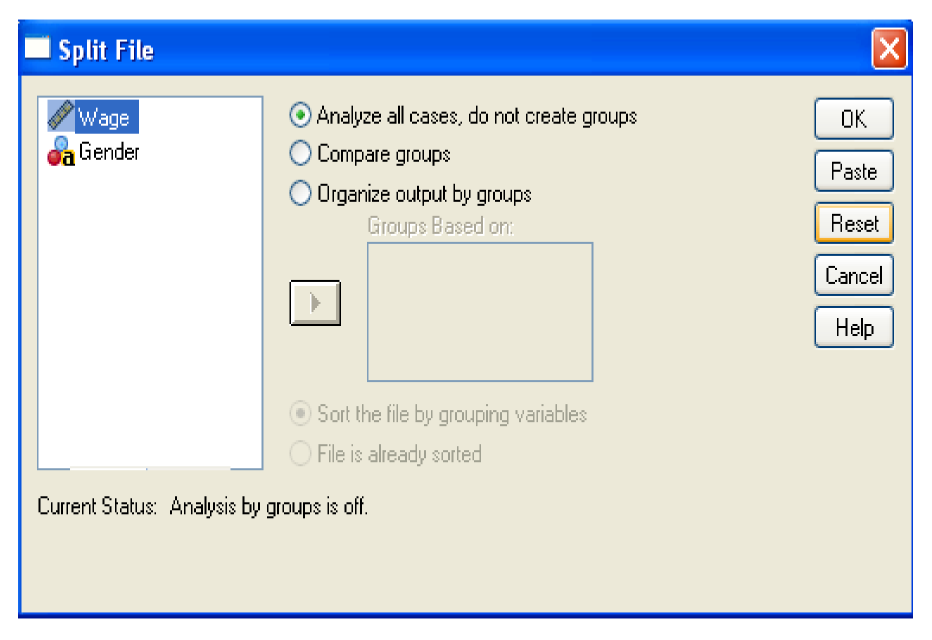
الجدول التالي يمثل رواتب مجموعة من الأشخاص حسب الجنس.

جدول (٧)

Wage	Gender
60	M
30	F
70	M
35	F
65	M
40	F

لتجزئة الملف إلى جزأين الأول يمثل رواتب الذكور M والثاني يمثل رواتب الإناث F نتبع الخطوات التالية :

- من شريط القوائم اختر Data ثم Split File سوف يظهر صندوق Split File (انظر الشكل ٢٥) .



الشكل (٢٥)

الصندوق يحتوي على جميع المتغيرات الموجودة في الملف وهناك عدة اختيارات وهي:

- ⌖ Analyze Cases،Do Not Create Groups عدم تجزئة الملف .
- ⌖ Compare Groups تجزئة الملف حسب فئات متغير معين .
- ⌖ Organize output by Groups ترتيب بيانات الملف بالنسبة لمتغير معين .

يوجد بعد ذلك خيارين مطلوب اختيار أحدهما (هذه الخيارات غير نشطة في حالة التعامل مع بيانات الملف دون تقسيم) وهما:

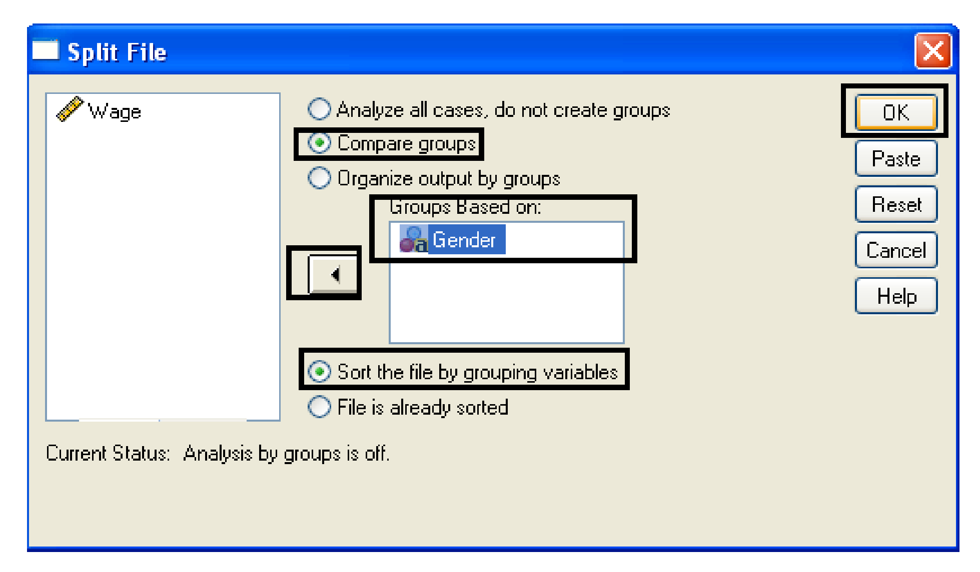
أ - ترتيب الملف طبقا للمتغير الذي تم على أساسه الفصل .

ب - لا داعي لترتيب المتغيرات في الملف .

ملاحظة : الخيارين أ ، ب تكون نشطة عند اختيار الخيارين ٢ ، ٣ .

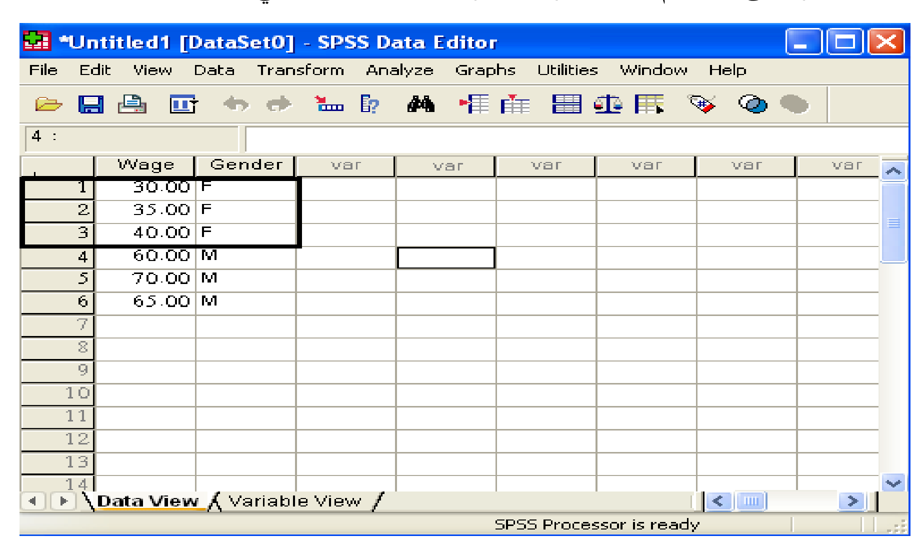
انقر على الخيار الثاني وانقل المتغير الذي على أساسه سيتم التقسيم وهو Gender إلى الخانة المخصصة لذلك وهي Grouped Based On .

- ⌖ انقر على أي من الخيارين وهما الأول لترتيب الملف طبقا لمتغير التقسيم Sort The File By Grouping Variables والثاني أن الملف غير مرتب Sort the file by grouping variables (انظر شكل ٢٦).



الشكل (٢٦)

⌖ انقر على Ok يتم تنفيذ الأمر فيظهر ملف النتيجة كما يلي:



الشكل (٢٧)

ترجيح حالات Weight Cases:

يتيح هذا الأمر إمكانية إعطاء أوزان لحالات Cases ملف معين نظرا لاختلافها من ناحية الأهمية و هي نفس الوظيفة التي تقوم بها الأيقونة ⚖ في شريط الأدوات .

مثال (٧):

القيم التالية تمثل نتائج أحد الطلاب في اختبار مادة الإحصاء علما أن لكل اختبار وزنا (أهمية) مختلفة:

جدول (٨)

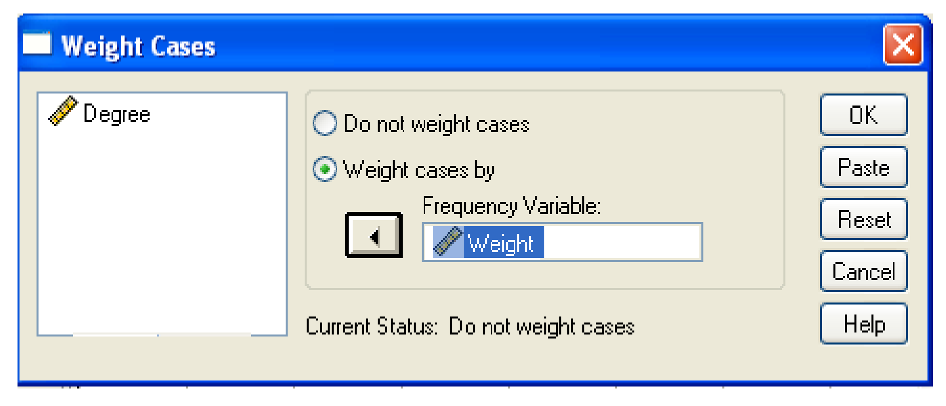
Test	Degree	Weight
A	100	50
B	60	35
C	75	5
D	55	15

المطلوب حساب الوسط الحسابي المرجح للامتحانات الأربعة؟

الحل:

لحساب الوسط الحسابي المرجح بالمتغير Weight للامتحانات الأربعة نتبع الخطوات التالية:

- من شريط القوائم Data نختار Weight Cases الذي نقوم بترتيبه (الشكل ٢٨):



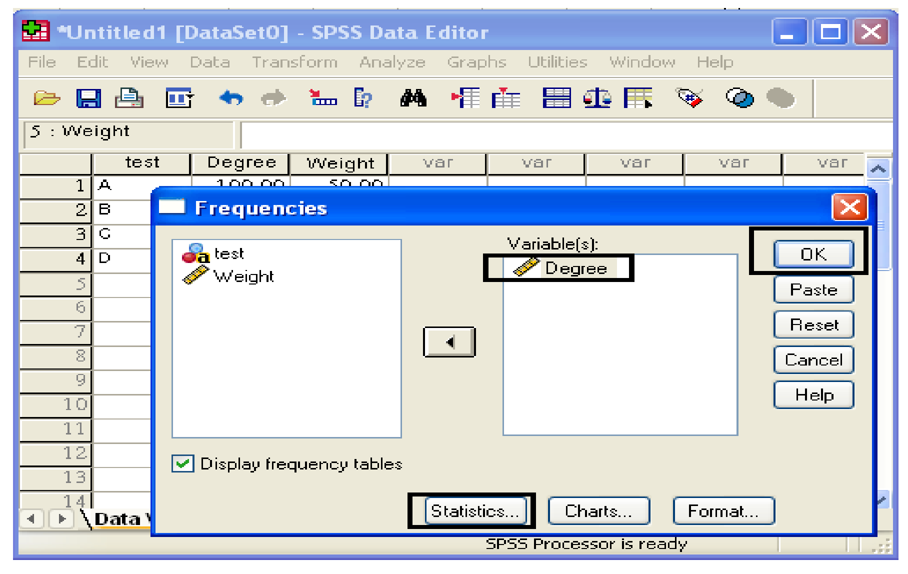
الشكل(٢٨)

- عند نقر زر Ok يتم وزن حالات الملف بالمتغير Weight (علما أنه لا يلاحظ أي تغير في شاشة Data Editor).

لحساب الوسط الحسابي نختار من شريط القوائم :

Analyze ➡ Descriptive Statistics ➡ Frequencies

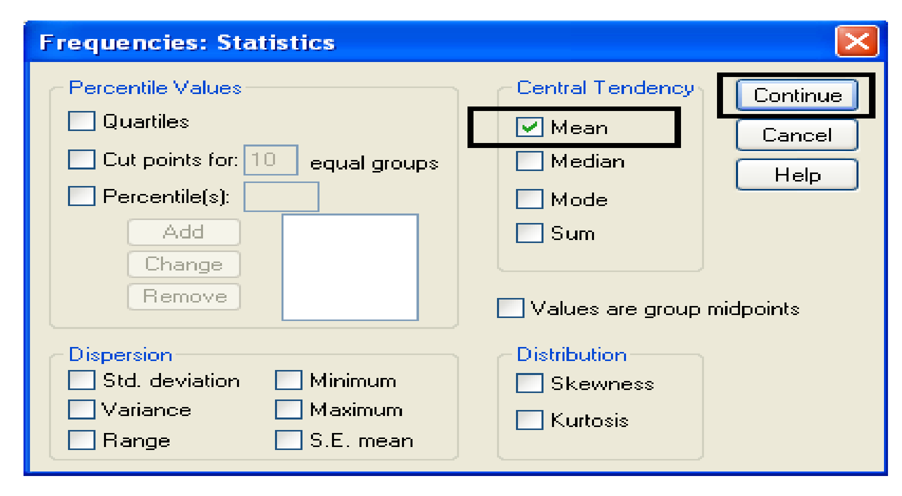
انظر الشكل (٢٩):



الشكل (٢٩)

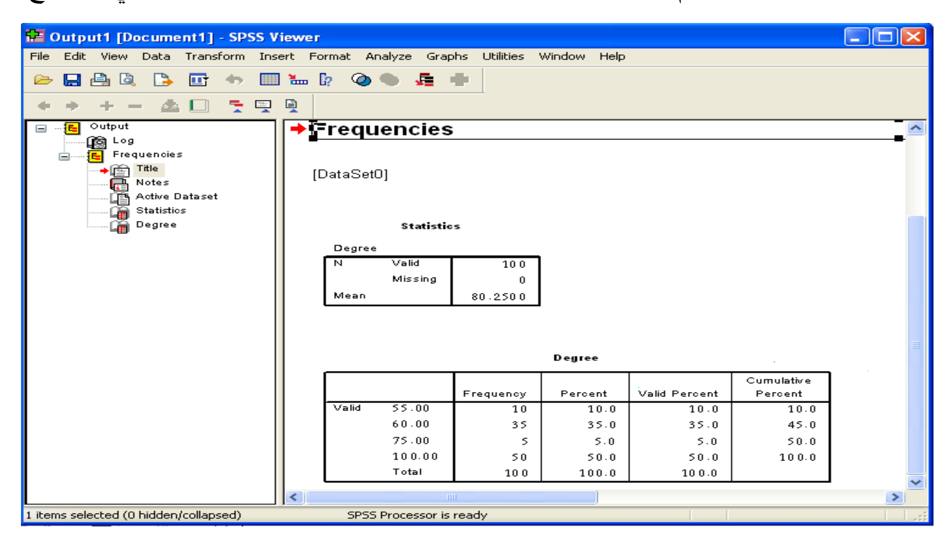
- نقوم بتحديد المتغير المراد حساب الوسط الحسابي المرجع لـه ويتم نقله عن طريق زر النقل ▶ ويمكننـا الرجـوع عـن الاختيـار باسـتخدام زر الرجـوع ◀

وباختيار Statistics نقوم بتأشير الخيار Mean كما في الشكل الآتي:



الشكل (٣٠)

⌖ وبنقر continue ثم Ok تظهر شاشة المخرجات الآتية بها الوسط الحسابي المرجح:



Statistics

Degree

N	Valid	100
	Missing	0
Mean		80.2500

Degree

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	55.00	10	10.0	10.0	10.0
	60.00	35	35.0	35.0	45.0
	75.00	5	5.0	5.0	50.0
	100.00	50	50.0	50.0	100.0
	Total	100	100.0	100.0	

الشكل (٣١)

ملاحظة هامة:

✓ لإلغاء ترجيح حالات الملف نقوم بتأشير الخيار Do Not Weight Cases في صندوق حوار Weight Cases.

الفصل الثالث

إنشاء متغيرات جديدة

Create New Variables

يمكن إجراء تحويلات على البيانات الأصلية ما بين تجميع بيانات في فئات معينة إلى تكوين متغيرات جديدة بالاعتماد على معادلات وصيغ شرطية فمثلا إذا كان لدينا بيانات استهلاك ودخول أفراد مجتمع معين فيمكننا من حساب حجم الادخار في هذا المجتمع عن طريق إنشاء متغير جديد عبارة عن الفرق بين الدخل والاستهلاك ليمثل الادخار تمهيداً لاستعماله في حساب حجم الاستثمارات في المجتمع، والبيانات التي لا تتبع التوزيع الطبيعي يمكن تحويلها عن طريق هذا الفصل إلى بيانات طبيعية. ونجد أنه في بعض الأحيان يصعب التعامل مع البيانات اللفظية ويكون التعامل أسهل عند تكويد البيانات اللفظية وتحويلها إلى أكواد رقمية ويتم ذلك عن طريق التحويلات. هذا الباب سيناقش كيفية التعامل مع القائمة Transformations.

قائمة Transform:

تضم قائمة Transform الأوامر التالية:

شكل (١)

استخدام العمليات الرياضية لإنشاء متغيرات مستمرة جديدة/الأمر Compute:

يتيح هذا الأمر إمكانية حساب متغيرات جديدة باستخدام أكثر من دالة تتضمن (دوال إحصائية ، توزيعات احتمالية)، أو تحويل متغيرات موجودة إلى متغيرات أخرى.

⌖ العمليات البسيطة:

يمكننا من استعمال بعض الدوال البسيطة مثل الجمع والطرح والضرب والقسمة والأس وهكذا في إجراء تحويل المتغيرات كما في المثال التالي.

مثال (١):

نفترض أننا نريد حساب حجم الادخار للبيانات التالية الموجودة في Data view التالي:

*your Analysis.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

8 : الاستهلاك 500 Visible: 2

	الدخل	الاستهلاك	var	var	var	var	var	var
1	50.00	23.00						
2	100.00	55.00						
3	150.00	70.00						
4	200.00	120.00						
5	300.00	200.00						
6	400.00	320.00						
7	500.00	450.00						
8	600.00	500.00						
9								
10								
11								
12								
13								
14								
15								

Data View / Variable View

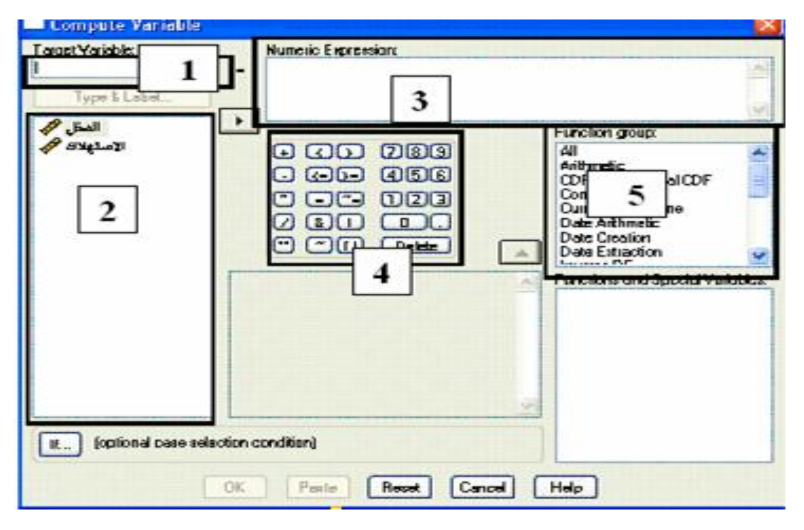
SPSS Processor is ready

شكل (٢)

فنجد أن: الادخار=الدخل - الاستهلاك

للحصول على عمود الادخار فإننا نقوم بطرح عمود الاستهلاك من عمود الدخل باستعمال الخطوات التالية:

⌖ من قائمة Transform نختار Compute Variable فيظهر الشكل الآتي:



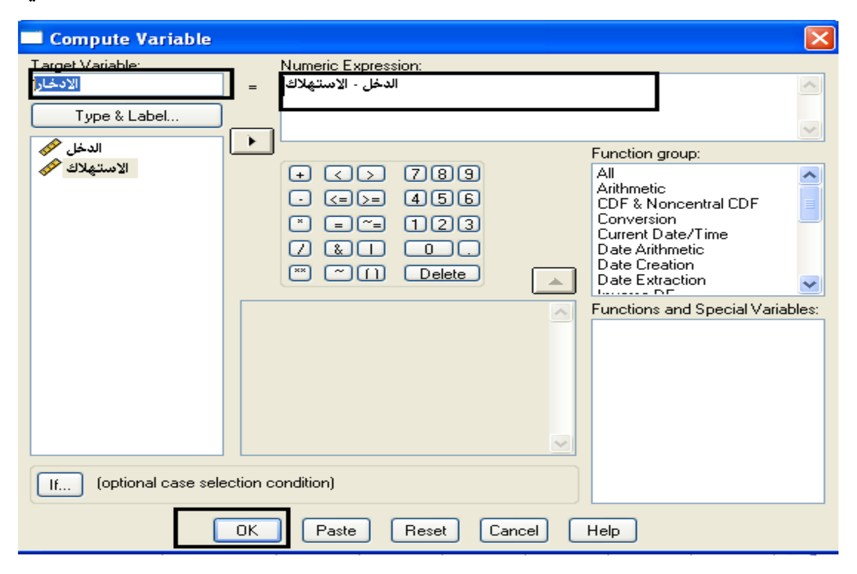
شكل (٣)

- في المنطقة رقم "١" نضع اسم المتغير الجديد.
- في المنطقة رقم "٢" وهي المتغيرات المتاح الاختيار منها لإجراء العمليات الحسابية عليها وإنشاء متغير منها ونلاحظ أننا لا نمتلك سوا متغيرين اثنين وهو الاستهلاك والدخل وظاهرين في الشكل (٣).
- في المنطقة رقم "٣" وهي المنطقة الذي يظهر بها شكل العمليات الحسابية المطلوبة.
- في المنطقة رقم "٤" هي منطقة تتيح لوحة مفاتيح للأرقام وبعض العمليات الحسابية البسيطة. والجدول التالي يوضح ملخص لبعض العمليات الحسابية البسيطة المتوفرة في هذة المنطقة:

جدول (١)

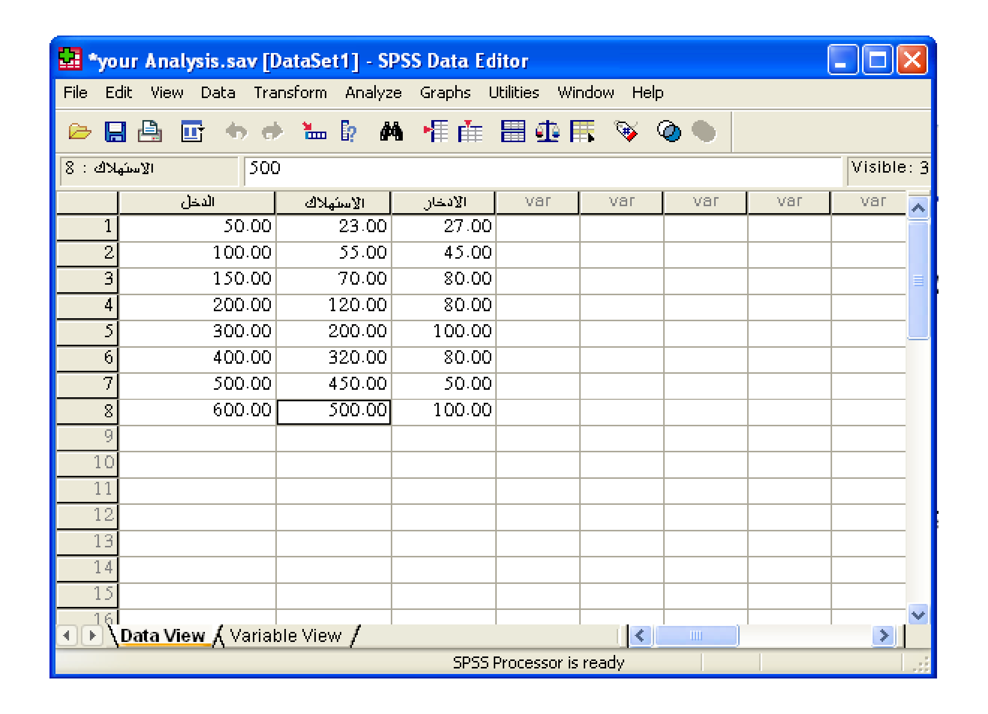
العملية الحسابية	الرمز
الجمع	+
الطرح	-
القسمة	/
الضرب	*
الأس	"**" أو "^"

- في المنطقة رقم "٥" هي منطقة الدوال الأكثر تعقيداً من الدوال الموجودة في المنطقة رقم "٤".
- ففي هذا المثال نقوم بكتابة كلمة "ادخار في المنطقة رقم "١" ليكون اسم المتغير الجديد.
- وبعدها نختار المتغير الدخل من المنطقة رقم "٢" عن طريق تحديد المتغير والنقر علي زر الانتقال [▸] ونلاحظ أن متغير الدخل ظهر بعد ذلك في المنطقة رقم "٣" تمهيداً لإجراء العمليات الحسابية عليه.
- وبعدها نختار علامة الطرح من المنطقة رقم ٤ وهي بالشكل الآتي [-] ونلاحظ أن بعد الاختيار فإن تلك العلامة تظهر في المنطقة رقم "٣" بعد متغير الدخل.
- وبعدها نختار مرة أخرى متغير الاستهلاك من المنطقة "٢". فيظهر الشكل الآتي:



شكل (٤)

- ننقر على Ok فتظهر شاشة data view كالتالي:



شكل (٥)

- **استخدام دوال SPSS لإنشاء متغيرات جديدة:**

يوفر لنا البرنامج مجموعة من الدوال ليست فقط الرياضية ولكن دوال إحصائية ومنطقية أيضا. ففي المثال الآتي يوضح كيف يمكننا استخدام الدوال الإحصائية في إنشاء متغيرات جديدة:

مثال (٢):

الجدول التالي يمثل المتغيرين X_1، X_2 اللذين تم إدخالها إلى Data Editor كما يلي:

جدول (٢)

X_1	X_2
60	90
87	88
70	43
90	80
57	55

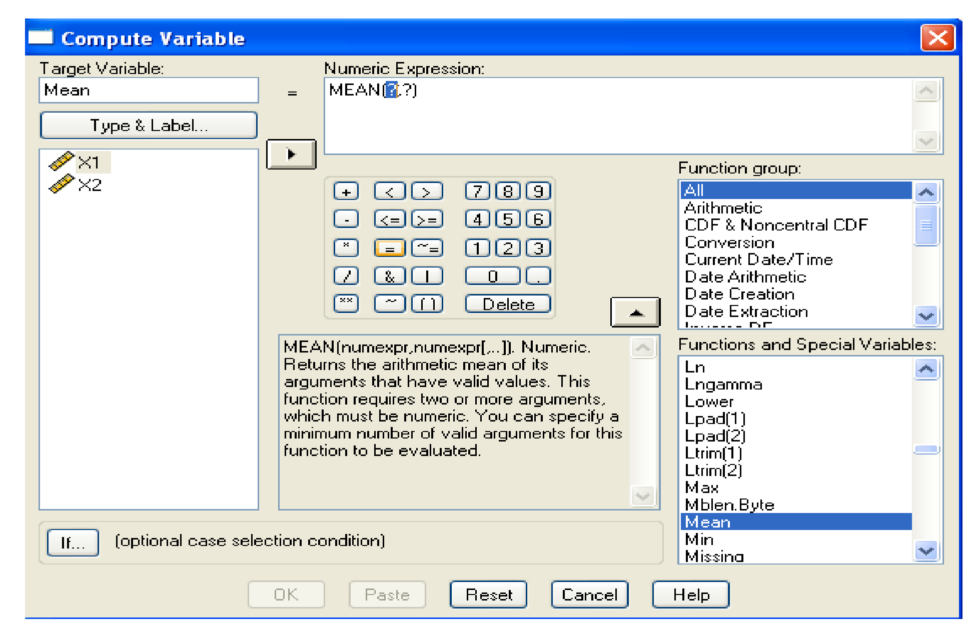
73	47
95	90
66	50
40	55
55	80
85	75
88	86
35	70

المطلوب حساب الوسط الحسابي Mean للمتغيرين X_1 و X_2 في حالة كون $X_1 \geq 50$، X_2.

الحل:

لتنفيذ ذلك نتبع الخطوات التالية:

- من شريط القوائم نختار Transform ثم ننقر Compute فيظهر صندوق حوار Comput Variable انظر الشكل (٦) ونكتب اسم المتغير الجديد وهو الوسط الحسابي (mean) لكلاً من X_1 and X_2 يلي ذلك اختيار الدالة Mean من منطقة الدوال المتاحة فتظهر في منطقة العمليات الحسابية وبجوارها علامتين استفهام العلامة الأولي متاحة لكتابة المتغير الأول والثاني متاحة للكتابة بدلاً منها المتغير الثاني ويمكن كتابة أكثر من متغير مفصول بين كل متغير والآخر بفصلة:



شكل (٦)

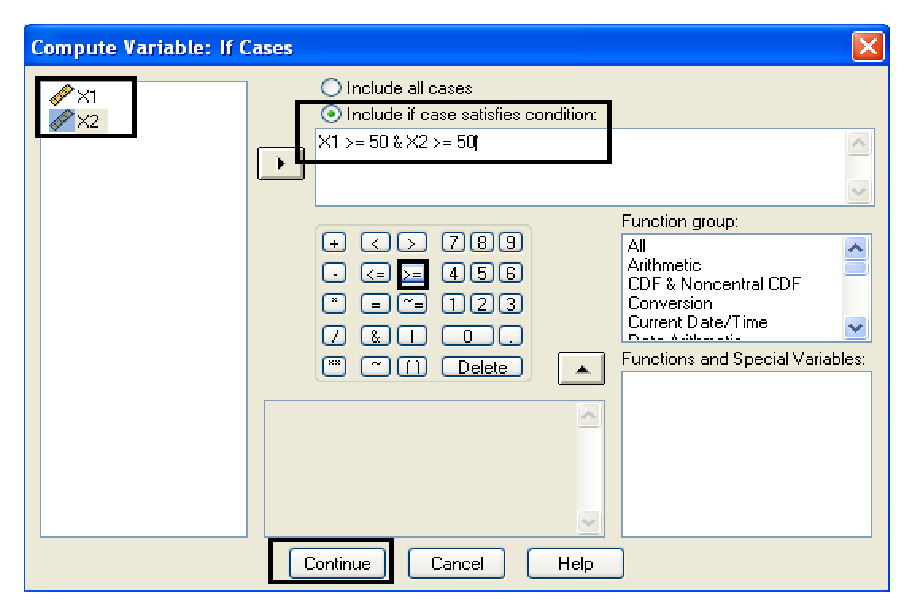
فنقوم بكتابة X_1 بدلاً من علامة الاستفهام الأولى "؟" وكتابة X_2 بدلاً من علامة الاستفهام الثانية "؟"

وقد أجرينا العمليات التالية:

- تحديد اسم المتغير Target Variable الذي هو عبارة عن الوسط الحسابي لـ X_1،X_2 و هو X_3.
- اختيار الدالة وهي Mean من قائمة Function.
- واضح في صندوق الحوار أعلاه.
- بعد الانتهاء انقر IF في صندوق حوار شكل (٦) حيث لدينا شرط في المتغير الجديد وهو أن يكون $X_1 \geq 50$،X_2 فيظهر صندوق حوار If Cases كما بالشكل (٧):

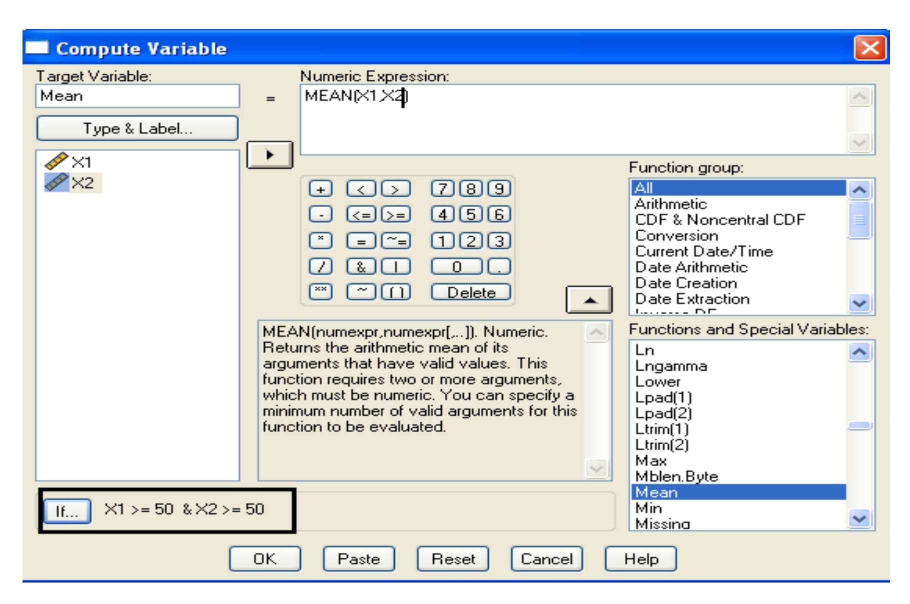
✓ لاختيار كافة الحالات انقر Include All Cases.

✓ لاختيار جزء من الحالات انقر Include If Cases Satisfies Condition. بما أننا نريد اختيار حالات نقوم بنقر الجزء الأخير. ونقوم بكتابة الشرط في شاشة الشرط كما في الشكل الآتي:



الشكل (٧)

⌖ انقر Continue في صندوق حوار if cases للعودة إلى صندوق الحوار الأصلي ونلاحظ أن الشرط قد تم كتابته بجوار أيقونة If كما في الشكل الآتي:



الشكل (٨)

- انفر زر ok في صندوق حوار Compute Variable.

يتم الحصول على نتائج حساب المتوسط حيث يتم إضـافة متغير آخر هو X_3 إلـى Data Editor كما يلي:

X1	X2	Mean
60.00	90.00	75.00
87.00	88.00	87.50
70.00	43.00	.
90.00	80.00	85.00
57.00	55.00	56.00
73.00	47.00	.
95.00	90.00	92.50
66.00	50.00	58.00
40.00	55.00	.
55.00	80.00	67.50
85.00	75.00	80.00
88.00	86.00	87.00
35.00	70.00	.

الشكل (٩)

الأمر Recode:

في كثير من الأحيان قد نحتاج إلى عملية التكويد (أو الترميز) وذلك بتعديل قيم البيانـات وتغييرهـا مـن بيانـات لفظيـة إلـى بيانـات رقميـة بغـرض تسـهيل التعامـل معهـا، ويستفاد منه أيضا عمل الفئات. يوجد عدة حالات يمكن فيها استخدام عمليـة التكويـد منهـا الآتـي:

- دمج بيانات كثيرة لمتغير مستمر في جـدول تكراري ليسـهل التعامـل معهـا. للتعـرف على تكويد البيـانات المستمرة نفرض أن لدينا المثال التالي :

مثال (٣):

لدينا المتغير Salary (X) (الأجر) يأخذ القيم التالية:

100, 150, 200, 250, 300, 500, 270, 300, 350, 700, 350, 400, 550, 600, 850

وقد تم ادخال قيم المتغير في شاشة Data Editor. وتم تخصيص كود لكل قيمـة مـن قيم المتغير حسب الترتيب التالي:

جدول (٣)

الفئة	الكود
١٥٠ فأقل	١
١٥١-٣٠٠	٢
٣٠١-٤٥٠	٣
٤٥١-٦٠٠	٤
٦٠١ فأكثر	٥

نرغب في ترميز المتغير Salary حسب الفئات المذكورة وخزن الرموز في متغير مختلف.

الحل:

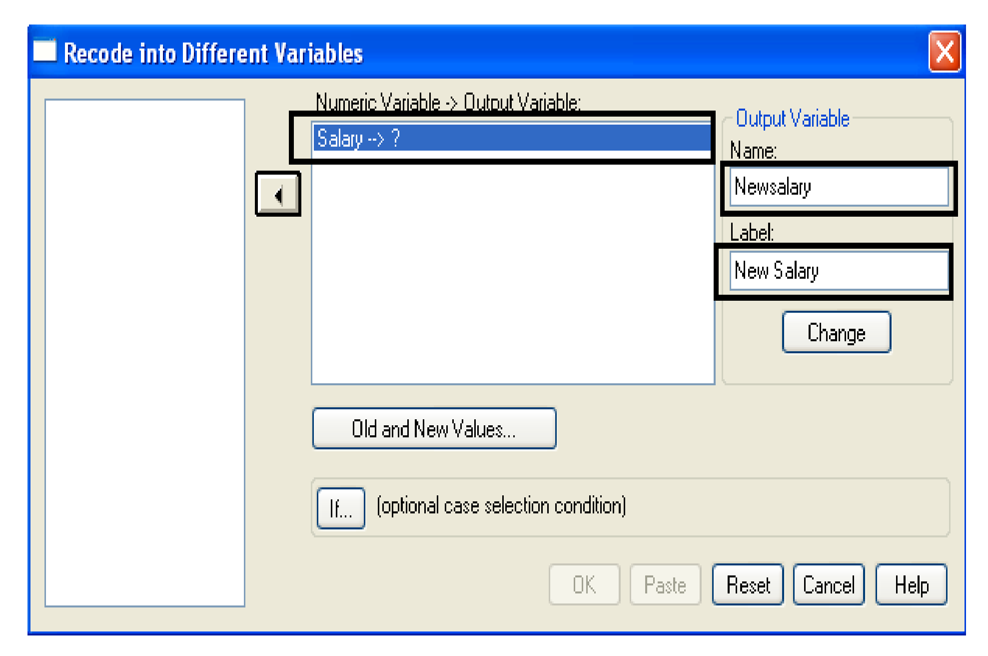
لتنفيذ ذلك نتبع الخطوات التالية :

- من القائمة Transforms عند النقر عليها سوف تظهر شاشة (١٠) وهنا يوجد في هذا المجال اختيارين:

* Recode into the Same Variable وهذا يعني أن نجري العملية ونضع القيم الجديدة فوق القيم القديمة للمتغير .

*Recode into Different Variable بينما يعني هذا أن نضع القيم الجديدة تحت اسم متغير جديد.

- ✓ عند الاختيار الأول تتم العملية تلقائياً، أما عند الثاني فسيطلب البرنامج معلومات عن المتغير الجديد.
- ✓ عند اختيار الأمر الثاني سيفتح الصندوق الحواري Recode Into Different Variable ننقل المتغير المراد تكويده إلى Input Variable وليكن Salary (الأجر)، ثم نحدد اسم المتغير الجديد في المكان المسمى Name وليكن Y انظر الصندوق الحواري الشكل (١٠):

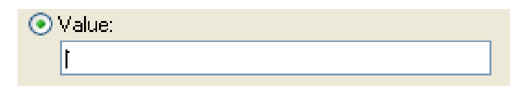


الشكل (١٠)

⌖ تبدأ عملية التكويد بالنقر على Old and new values فيفتح الصندوق الحواري

Recode Into Different Variable: Old And New Values انظر الشكل (١١):

✓ فإذا كنا نريد تكويد قيمة



فإننا نقوم بنقر تحت الأمر Old Values ولكننا هنا نريد تكويد فئات لها أكثر من نوع.

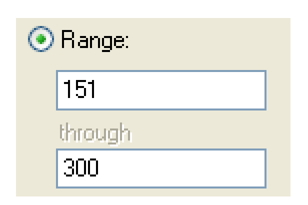
✓ أول فئة لدينا هي ١٥٠فأقل فإننا نختار



تحت الأمر Old Values ثم يلي ذلك أن ندخل الترميز "١" في المستطيل الموجود تحت كلمة New value ويكون بشكل الآتي :

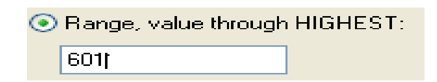
Value: 1

- ثم ننقر كلمة Add فتنتقل القيم المختارة إلى الصندوق Old – New الفئات التالية حتى الفئة قبل الأخيرة هي فئة لها حد أدنى وحد أعلى معرف "١٥١-٣٠٠" والحد الادني هو ١٥١ والحد الأعلى هو ٣٠٠ وهكذا باقي الفئات ففي هذه الحالة نقوم فإدخال الحد الأدنى في الخانة العلوية والحد الأعلى في الخانة السفلية كما يلي:



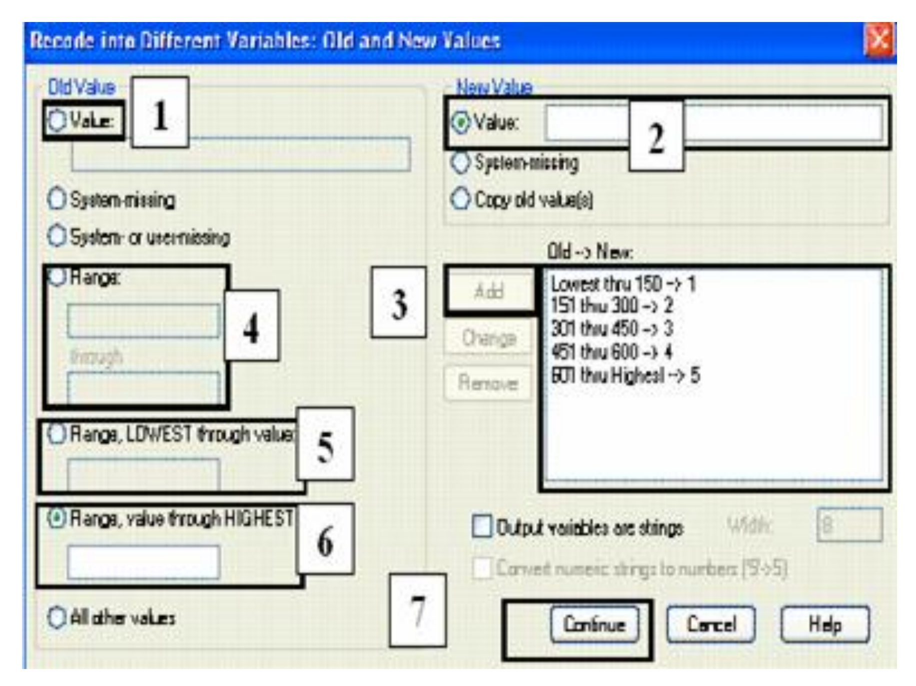
ثم يلي ذلك أن ندخل الترميز "٢" في المستطيل الموجود تحت كلمة New value ثم النقر علي add لانتقال إلي إدخال فئة جديدة لتكويد.

- الفئة الأخيرة لها حد ادنى وليس لها حد أعلى وتسمي الفئة المفتوحة من أسفل فنقوم بإدخال الحد الأدنى لها كما يلي:



وبعد ذلك ثم يلي ذلك أن ندخل الترميز "٥" في المستطيل الموجود تحت كلمة New value

- عند الانتهاء من عملية التكويد يظهر الشكل (١١) التالي، في النهاية ثم نختار الأمر Continue لنعود إلى صندوق الحوار الأصلي.



الشكل (١١)

⌖ ننقر Change ليتم التغيير من قيم المتغير إلى الترميز الجديد.

⌖ ننقر على الأمر Ok فيتم تغيير القيم الأصلية إلى الترميز الجديد حسب الفئات تحت اسم متغير جديد يرمز له بالاسم Newsalary كالآتي:

Salary	Newsalary
100.00	1.00
150.00	1.00
200.00	2.00
250.00	2.00
300.00	2.00
500.00	4.00
270.00	2.00
300.00	2.00
350.00	3.00
700.00	5.00
350.00	3.00
400.00	3.00
550.00	4.00
850.00	5.00

تكويد متغير لفظي إلى متغير رقمي:

- **إنشاء متغيرات وهمية و فئوية باستخدام التكويد**

يعتبر التكويد أداة هامة في تحليل العلوم الاجتماعية حيث تكون عملية المقارنة بين المجموعات المختلفة عملية هامة في التحليل مثل المقارنة بين كفاءة الذكور و الإناث في رياضة معينة، المقارنة بين جودة التعليم العامة والخاصة في الجامعات ... الخ. وعلى ذلك فإنه ينبغي لتحليل تلك المقارنات استخدام متغيرات وهمية و متغيرات فئوية.

- **المتغيرات الوهمية والمتغيرات الفئوية:**

المتغيرات الوهمية هي المتغيرات التي تأخذ قيمتين عادة ما تكون صفر و واحد. كل قيمة ترمز إلى فئة معينة والقيمة الأخرى ترمز إلى الفئة الثانية.

جدول (٤)

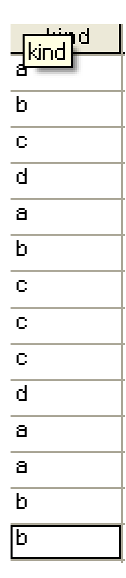
Value	Category
0	Male
1	Female

- المتغيرات الفئوية هي المتغيرات التي تحتوي على أكثر من قيمتين كل قيمة ترمز إلى فئة معينة ولا توجد أفضلية لفئة عن الأخرى مثل أصول الجنسية في الولايات المتحدة تنقسم إلى ٦ فئات والتحويل من الفئات إلى القيمة موضح في الجدول التالي:

جدول (٥)

Value	Category
0	White-American
1	African-American
2	Asian-American
3	Hispanic-American
4	Native-American
5	Other

و الآن نقوم بشرح كيفية تحويل تلك المتغيرات باستخدام البرنامج. نفترض أننا نريد التكويد إلى متغير جديد وهي متغير موصوف كما يلي:



لتنفيذ ذلك نمر بكل الخطوات السابقة إلى أن ينفتح الصندوق الحواري Old And New Values من الأمر Old Values نختار أول مستطيل يسارا من أعلى المسمى Value ويكتب به الحرف المراد تكويده ثم نضع الرقم المناظر لذلك الحرف في أول مستطيل يمينا من أعلى New Value وننقر Add لتضاف إلى القائمة Old-New، نكرر هذه العملية لنحصل على الصندوق الحواري الشكل (١٢):

الشكل(١٢)

ننقر Continue ونمر بنفس الخطوات السابقة إلى أن يتم التغيير فيصبح المتغير بعد التكويد بالشكل الآتي:

kind	newkind
a	1.00
b	2.00
c	3.00
d	4.00
a	1.00
b	2.00
c	3.00
c	3.00
c	3.00
d	4.00
a	1.00
a	1.00
b	2.00
b	2.00

إحلال قيم مفقودة بقيم أخرى Replacing Missing Values :

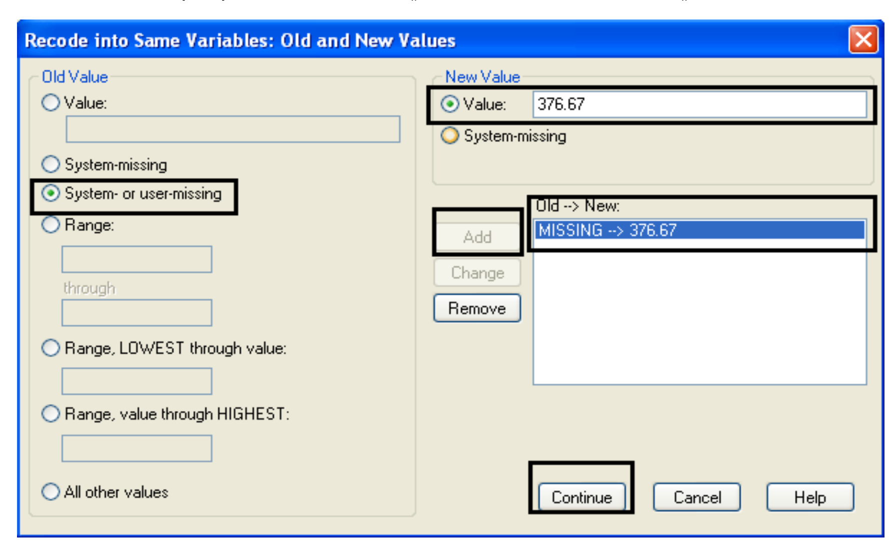
إن وجود قيم مفقودة لبعض المتغيرات تعتبر أحيانا عقبة كبيرة تواجه تطبيق أسلوب إحصائي معين ويتوجب في هذه الحالة تقدير القيم المفقودة حيث يوفر البرنامج SPSS هذه الطريقة. بفرض أن متغير الدخل به مجموعة من القيم المفقودة، تتم عملية الاستبدال عن طريق البرنامج كما يلي:

من القائمة Transform انقر الأمر Recode into same variables فينفتح الصندوق الحواري. اختر المتغير (أو المتغيرات)من قائمة المتغيرات وانقلها إلى المستطيل Variables.

- انقر على Old And New Values لفتح الصندوق الحواري التالي الخاص بعمليات التكويد والمسمى Recode into same Values : Old And New Values.
- في المربع Old Value نختار System-Or User –Missing.
- في المستطيل New Value اكتب القيمة المراد إحلالها مكان القيمة المفقودة ولتكن الوسط الحسابي للبيانات مثلا. ننقر الأمر Add فيتم الاختيار.

⌖ انقر الأمر Continue للعودة للصندوق الأصلي.

انظر الصندوق الآتي بفرض أن الوسط الحسابي 376.67 الشكل (١٣).



الشكل (١٣)

⌖ ننقر الأمر Ok لتنفيذ العملية.

الترتيب Ranking:

يمكن بواسطة هذا الأمر تكوين متغيرات جديدة هي عبارة عن رتب متغيرات معينة وتكون هذه الرتب تصاعدية أو تنازلية ويمكن اعطاء رتب لمتغير بواسطة متغيرات أخرى ، حيث أننا في كثير من طرق العرض والتحليل الإحصائي نرغب في التعامل مع الرتب الخاصة بالمتغير وليس بقيم المتغير.

مثال (٣):

لنفترض أن لدينا المتغيرات التالية: الدخل، الاستهلاك والادخار كالتالي:

جدول (٦)

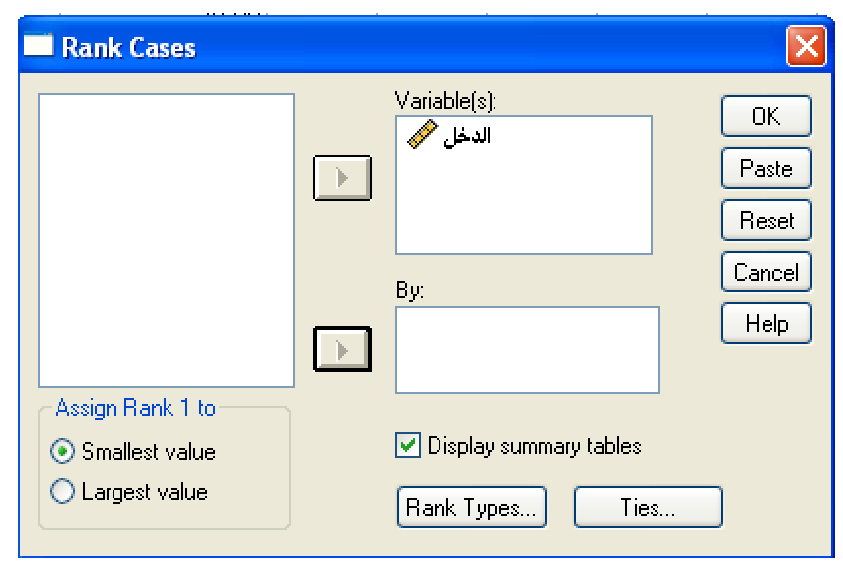
الاسم	الدخل
احمد	٥٠.٠٠
محمود	١٠٠.٠٠
مني	١٥٠.٠٠
هبة	١٥٠.٠٠
امال	٢٠٠.٠٠
محمد	٣٠٠.٠٠
شيماء	٤٠٠.٠٠
عز الدين	٥٠٠.٠٠
نورا	٥٠٠.٠٠
عبد التواب	٦٠٠.٠٠
عمرو	٦٣٠.٠٠
مهدي	٨٣٠.٠٠

ونريد إعطاء رتبا تصاعدية لمتغير الدخل .

الحل:

لتنفيذ ذلك نتبع الخطوات التالية:

- من القائمة Transform نختار Rank Cases فيظهر صندوق حوار Rank Cases الذي نرتبه كما في الشكل (١٤):



الشكل (١٤)

- لقد قمنا بإدخال العمود الذي نريد إعطاؤه رتبة الدخل في خانة Variables
- وفي خانةAssign Rank To أشرنا الخيار Smallest Value لإعطاء الرتب تصاعديا.
- لاختيار نوع الرتبة انقر زر Rank Types في صندوق حوار Rank Cases يظهر صندوق حوار Rank Cases: Types فيظهر أنواع مختلفة من الرتب:
- Rank: وهي الرتبة البسيطة (حيث يتم إعطاء رتبة لكل قيمة من قيم المتغير تعبر عن ترتيبه ضمن المجموعة).
- Savage Scores: تعطى رتب لقيم المتغير بموجب التوزيع الآسي.
- Fractional Rank: وهي الرتبة الناتجة من قسمة الرتبة البسيطة لقيم المتغير على مجموع الأوزان لكافة الحالات (أو عدد الحالات في حالة عدم وجود أوزان أي يعتبر الوزن مساوياً الي الواحد).
- Fractional Rank as %: يتم الحصول على هذة الرتبة من حاصل ضرب الرتبة السابقة في ١٠٠.
- Sum of cases weights : الرتبة تكون متساوية لكافة الحالات وتمثل مجموع الأوزان لكافة الحالات (أو عدد الحالات في حالة عدم وجود أوزان).

⌖ Ntiles: يتم إعطاء رتب بعد تقسيم قيم المتغير الى مجاميع تعطي كل منها رتبة معينة (بعد ترتيبها تصاعدياً أو تنازلياً) فإذا اخترنا 4Ntiles (حالة الترتيب تصاعدي) فإن سيتم إعطاء الرتبة ١ للقيم التي ترتيبها أقل من 25% وتعطى الرتبة 2 للقيم من 25% إلى 50% و الرتبة 3 للقيم 50% إلى 75% و الرتبة 4 لـ75% فما فوق.

⌖ و هنا في هذا المثال يتم اختيار النوع البسيط للترتيب Rank انظر الشكل(١٥):

Rank Cases: Types

Rank
Savage score
Fractional rank
Fractional rank as %
Sum of case weights
Ntiles: 4
Continue
Cancel
Help
Proportion estimates
Normal scores
Proportion Estimation Formula
Blom Tukey Rankit Van der Waerden

الشكل(١٥)

⌖ ثم ننقر على Continue

النقر على الأمر Ties يعطي الصندوق الحواري الفرعي Ties: Rank Cases أنظر الشكل (١٦):

Rank Cases: Ties

Rank Assigned to Ties
Mean Low High
Sequential ranks to unique values
Continue
Cancel
Help

الشكل (١٦)

وفيه يختار المستخدم طريقة التعامل مع التداخلات Rank Assign to Ties هل باستخدام الوسط أم القيمة الصغرى أم الكبرى للرتب المتداخلة.

⌖ ثم زر Ok.

فيضاف متغير جديد (متغير الرتب) باسم الدخل R إلىData Editor كما في الجدول التالي:

جدول (٧)

الاسم	الدخل	الدخلR
احمد	50.00	1.000
محمود	100.00	2.000
منى	150.00	3.500
هبة	150.00	3.500
امال	200.00	5.000
محمد	300.00	6.000
شيماء	400.00	7.000
عز الدين	500.00	8.500
نورا	500.00	8.500
عبد الثواب	600.00	10.000
عمرو	630.00	11.000
مهدي	830.00	12.000

إنشاء سلسلة زمنية Create Time Series:

هي عبارة عن قيم متغير معين خلال فترات زمنية متساوية كالأيام أو الأشهر أو السنوات، ويمكن اجراء بعض العمليات الإحصائية على السلسلة الزمنية من خلال عدة دوال إحصائية تضمن:الفروق Differences، الأوساط المتحركة Moving Averages، التأخر Lag، التقدم Lead، وغيرها.

مثال (٤):

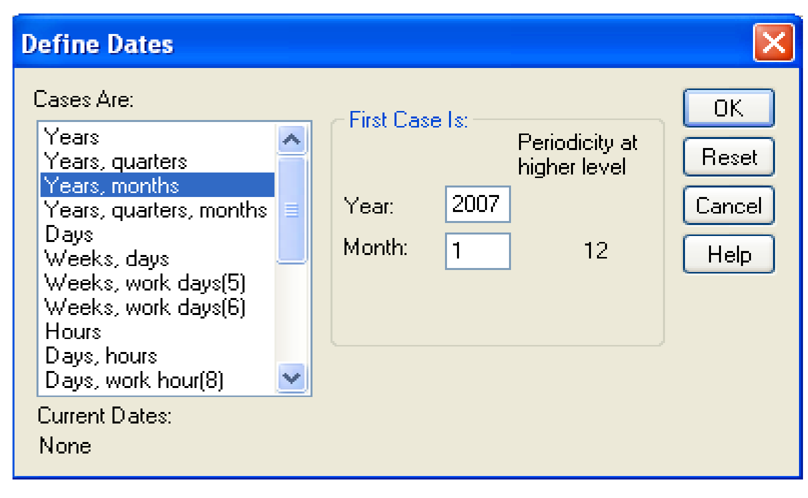
لدينا المتغير حجم المبيعات ويمثل المبيعات الشهرية من أجهزة المحمول خلال 12 شهراً في مؤسسة معينة لسنة ٢٠٠٧ م ونرغب في عمل فروق Differences من الدرجة الأولى لهذا المتغير:

274 207 255 350 382 383 351 268 380 409 445 455

الحل:

- أولاً نقوم بإنشاء السلسلة الزمنية بإنشاء الشهور بالنسبة لسنة 2007 كالتالي:

انقر على الأمر Define Dates من القائمة Data فيظهر صندوق الحوار Define dates الشكل(١٧)الذي نقوم بترتيبه كما يلي:



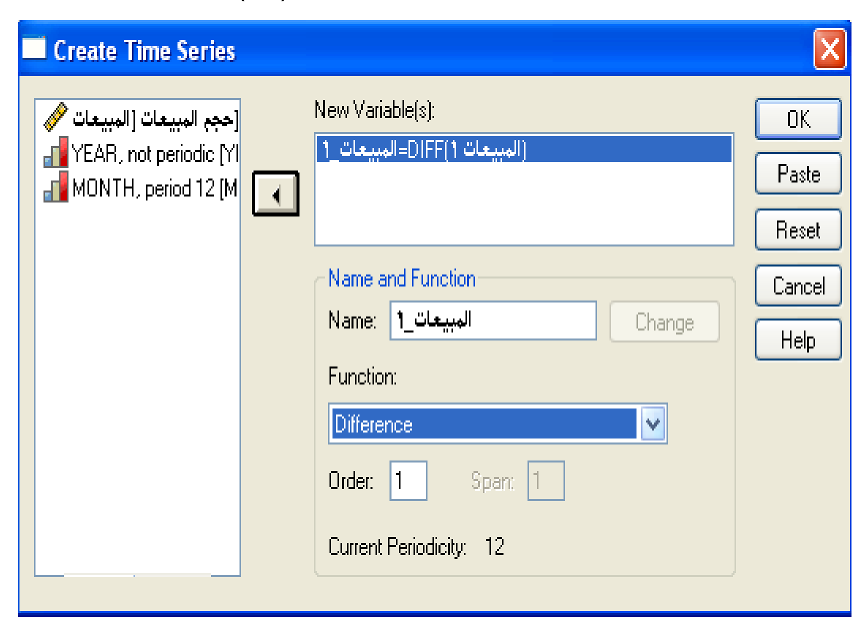
الشكل (١٧)

حيث اخترنا من Cases are السنة والشهور Years،Months ويمكن للباحث أن يختار ما يشاء سواء الأيام أو الساعات أو الأسابيع وغيرها.

- في قائمة First case is نقوم بتعريف التاريخ للحالة الأولى في السلسلة الزمنية: في حقل السنة ندخل سنة البداية 2007، أما في حقل الشهر ندخل شهر البداية 1.
- أماPeriodicity at Higher Level فيبين دورية التاريخ حيث يبين أكبر قيمة يمكن تزويدها للبرنامج مثلاً في المثال أعلى دورية للشهور 12 شهراً.
- عند نقر Ok في صندوق حوار Define Dates تضاف متغيرات التاريخ إلى Data Editor كما يلي:

المبيعات	YEAR_	MONTH_	DATE_
274.00	2007	1	JAN 2007
207.00	2007	2	FEB 2007
255.00	2007	3	MAR 2007
350.00	2007	4	APR 2007
382.00	2007	5	MAY 2007
383.00	2007	6	JUN 2007
351.00	2007	7	JUL 2007
268.00	2007	8	AUG 2007
380.00	2007	9	SEP 2007
409.00	2007	10	OCT 2007
445.00	2007	11	NOV 2007
455.00	2007	12	DEC 2007

- ثانياً نقوم بعمل الفروقات Differences بأتباع الخطوات التالية :
- انقر على الأمر Create Time Series من القائمة Transform، يظهر الصندوق الحواري Create Time Series حدد اسم المتغير الذي تعتمد عليه السلسلة وانقله إلى المستطيل New Variables أنظر الشكل (١٨).



الشكل (١٨)

حدد اسم المتغير الجديد في المستطيل Name تحت الأمر Name and Function وليكن" المبيعات _١". يتم تحديد الدالة الجديدة التي سيتم على أساسها إنشاء بيانات

المتغيـر الجديـد مـن الأمـر Function والتــي فــي المثــال هــي Differences، المتغيـر الجديد المختار هو دالة الفروق.

عند نقر Ok يضاف متغير جديد باسم المبيعات _١ إلى Data Editor كم يلي في الشكل (١٩):

المبيعات	YEAR_	MONTH_	DATE_	المبيعات_١
274.00	2007	1	JAN 2007	.
207.00	2007	2	FEB 2007	-67.00
255.00	2007	3	MAR 2007	48.00
350.00	2007	4	APR 2007	95.00
382.00	2007	5	MAY 2007	32.00
383.00	2007	6	JUN 2007	1.00
351.00	2007	7	JUL 2007	-32.00
268.00	2007	8	AUG 2007	-83.00
380.00	2007	9	SEP 2007	112.00
409.00	2007	10	OCT 2007	29.00
445.00	2007	11	NOV 2007	36.00
455.00	2007	12	DEC 2007	10.00

الشكل (١٩)

الفصـــل الرابع

اختبارات عينة أو عينتين

One and Two Samples Tests

تعتبر اختبارات الفروض الإحصائية واحدة من أهم التطبيقات التي قدمها علم الإحصاء كحل للمشاكل العلمية المختلفة بشتى فروع العلم .

وتبدأ أي مشكلة بحثية باهتمام الباحث بدراسة خصائص مجتمع الدراسة وتتم دراسة هذه الخصائص بعدة طرق منها اختبار فرض (أو فروض) يتعلق بها لكي يتأكد الباحث من صحة أو عدم صحة هذا الفرض حيث أنه لا يستطيع التعامل مع المجتمع ككل لان هذا إما مستحيلا أو مكلفا للغاية ولذلك ليس أمام الباحث إلا اختيار عينة عشوائية من ذلك المجتمع، ثم اختبار الفروض الموضوعة للبحث تهيئة لاستخلاص النتائج.

بصفة عامة وقبل الدخول في أية تفاصيل فإن اتخاذ أي قرار بقبول أو رفض فرضا معينا يمر بعدة مراحل وهي كالتالي:

- سحب عينة من المجتمع بحيث تكون ممثلة تمثيلاً جيداً للمجتمع.
- تجميع البيانات المتعلقة بالمشكلة من العينة مع توخي الدقة في التجميع حتى تكون النتائج قابلة للتطبيق بأقل الأخطاء الممكنة.
- تطبيق قواعد معينة لاختبار الفروض الموضوعة عن طريق الباحث وهي مشكلة تحتاج إلى خبرة إحصائية .

تعاريف أساسية في اختبارات الفروض الإحصائية

Testing Statistical Hypotheses:

فيما يلي نقدم مجموعة من المصطلحات التي تستخدم في اختبارات الفروض الإحصائية:

- **الفرض Hypothesis:**

هو تقرير أو تأكيد عن التوزيع الاحتمالي و قد يكون هذا التقرير أو التأكيد assertion متعلق بأحد معالم التوزيع الاحتمالي أو بأحد المؤشرات التي تعكس خاصية معينة من خواص التوزيع الاحتمالي.

- **الفرض العدمي H_0 (Null Hypothesis):**

فرض العدم (الفرض الصفري) نرمز له بالرمز H_0 وهو فرض حول معلمة المجتمع التي نجري الاختبار عليها باستخدام بيانات من عينة والتي تشير أن الفرق بين معلمة المجتمع والإحصاءه من العينة ناتج عن الصدفة ولا فرق حقيقي بينهما في حالة قبوله.

وهذا الفرض نرفضه عندما تتوفر لدينا دلائل على عدم صحته، وتعني كلمة Null أنه لا يوجد فرق بين معلمة المجتمع والقيمة المدعاة (إحصائية العينة).

- **الفرض البديل H_A (Alternative Hypothesis):**

هو الفرض الذي نرمز له بالرمز H_A ويضعه الباحث كبديل عن فرض العدم و نقبله عندما نرفض الفرض العدم باعتباره ليس صحيحاً بناءً على المعلومات المستقاة من العينة.

- **اختبار الفرض الإحصائي A test of a statistical Hypothesis**

يعرف اختبار الفرض الإحصائي بأنه القاعدة التي بناء على دراسة مفردات العينة تؤدي إلى قرار معين بشأن قبول أو رفض فرض العدم. ويلاحظ أن القرار الذي نصل إليه طبقا للقاعدة المشار إليها يتعلق عادة بفرض العدم أي H_0 ، فإما أن يكون القرار هو رفض H_0 (وفي هذه الحالة يعني ذلك قبول H_A) أو قبول H_0 (وفي هذه الحالة يعني ذلك رفض H_A).

- **الخطأ من النوع الأول و الخطأ من النوع الثاني**

Type 1 error and Type 2 error

يعتمد القرار الذي يتم اتخاذه بشأن فرض العدم على المعلومات التي تعطيها العينة العشوائية التي يتم سحبها. أي أن قبول فرض العدم (رفض الفرض البديل) أو رفض فرض العدم (قبول الفرض البديل) يعتمد على النتيجة التي تعطيها العينة العشوائية، وهنا قد تعطي العينة معلومات أو نتيجة تؤدي إلى رفض فرض العدم برغم كونه في الحقيقة صحيحا، وفي هذة الحالة يكون القرار خاطئا، ومن ناحية أخرى قد تعطي العينة معلومات أو نتيجة في صالح فرض العدم فيتم قبوله رغم كونه في الحقيقة غير صحيح وهذا قرار خاطىء آخر.

وفي الحقيقة بسبب اعتماد القرار الذي يتم اتخاذه بشان فرض العدم على نتيجة العينة العشوائية فإن الحالات الممكنة للقرار الذي يتخذه بشأن الفرض العدم هي كما يلي:

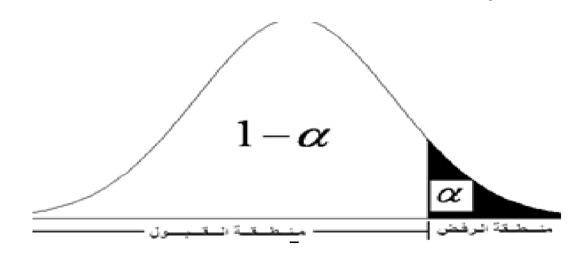
القــــرار		على مستوي المجتمع	
		H_0 صحيح	H_0 غير صحيح
على مستوى العينة	قبول H_0	القرار صحيح	خطا من النوع الثاني
	رفض H_0	خطا من النوع الأول	القرار صحيح

⌖ **أنواع اختبارات الفروض:**

عندما نقبل فرض العدم فإننا نقبله بمستوى دقة معين فقد تكون ٠.٩٠ أو ٠.٩٥ أو ٠.٩٩ أو غير ذلك وتسمى مستويات الثقة Significance Levels أي يوجد نسبة خطأ معين في قبولنا للفرض العدم ونرمز لها بالرمز α ويسمى مستوى المعنوية، أي إذا كان مستوى الثقة-1 α =() 0.95 فان مستوى المعنوية α تساوي ٠.٠٥ وهي عبارة عن مساحة منطقة تحت منحنى التوزيع تمثل منطقة الرفض وتكون إما على صورة ذيل واحد جهة اليمين أو اليسار أو ذيلين متساويين في المساحة واحد جهة اليمين والثاني جهة اليسار.

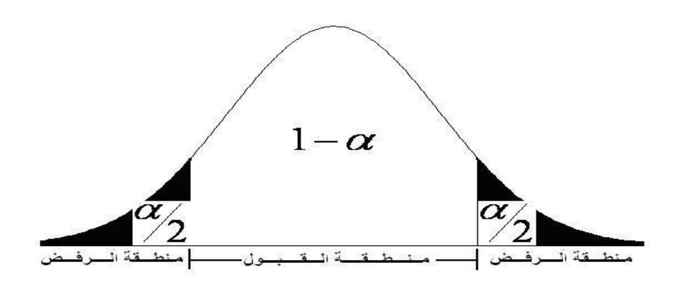
⌖ **تعريف اختبار الفروض في جانب واحد:**

هو الاختبار الذي يضع الفرض البديل للمعلمة المجتمع علي انها أكبر أو أصغر من قيمة فرض العدم لتلك المعلمة، فتكون منطقة الرفض إما من اليمين اذا كان الفرض البديل أكبر أو من اليسار إذا كان الفرض البديل أقل.



- **تعريف اختبار الفروض في جانبين(ذيلين):**

هو الاختبار الذي يكون فيه الفرض البديل عن معلمة المجتمع على هيئة $\neq$ ، بل مجرد أنها تختلف. اختبار من طرفين (من الجهة اليمنى و اليسرى) $H_A : \theta \neq \theta_0$



- **خطوات اختبارات الفروض الإحصائية لحجم ثابت للعينة**

Steps for Statistical Hypotheses Testing With Fixed Sample Size

عند إجراء اختبارات الفروض الإحصائية عملياً تكون هناك خطوات محددة لإجراء الاختبار حيث تكون لازمة وبالترتيب المحدد لكل منها لإجراء الاختبار واتخاذ القرار بشأنه بطريقة سليمة. والخطوات التالية تحدد كيفية إجراء اختبارات الفروض الإحصائية التي تعتمد على حجم ثابت للعينة:

- صياغة كل من الفرض العدم H_0 والفرض البديل H_A حسب المشكلة المراد إجراء اختبارات الفروض لها.
- تحديد مستوى المعنوية للاختبار بما يلائم تكلفة الوقوع في الخطا من النوع الأول أو الخطأ من النوع الثاني.
- تحديد دالة الاختبار وهي إحصائية أي أنها تعتمد فقط على مشاهدات العينة أو الفرض العدم وتختار بحيث يكون توزيعها الاحتمالي معلومة تماما تحت فرض العدم أي أن دالة الاختبار ما هي إلا دالة في مشاهدات العينة توزيعها الاحتمالي معلوم.
- تكوين قاعدة القرار وتعتمد على كل من الفرض البديل ومستوى المعنوية ونقصد بقاعدة القرار أي هي القاعدة التي يبنى عليها اتخاذ القرار إما بقبول فرض العدم إذا وقع خارج المنطقة الحرجة أي منطقة الرفض للاختبار) أو يتم اتخاذ قرار برفض فرض العدم (إذ وقعت داخل المنطقة الحرجة أي منطقة الرفض).

- حساب قيمة دالة الاختبار.
- اتخاذ القرار هناك طريقتين لاتخاذ قرار في الاختبارات الإحصائية وذلك إما برفض أو قبول الفرض العدمي:

الطريقة الأولى:

تعتمد على قيم محسوبة لدالة الاختبار وقيم جدولية للاختبار تستخرج من جداول أعدت خصيصا لذلك الغرض.

الطريقة الثانية:

لا تتطلب وجود جداول بل تعتمد على قيم تحسبها الحزمة SPSS أو أية حزمة إحصائية ونطلق عليها p-value ، ونستطيع تلخيص الطريقتين كالتالي :

الطريقة الأولى:

- الاختبار الإحصائي قد يكون من طرف واحد (يمين أو يسار) One Tail Test أو من طرفين Two Tail Test ، إذا كان الاختبار من طرفين نوجد قيمتين حرجتين نحصل عليهما من جداول التوزيع العيني الذي يتحكم في الاختبار أحدهما في الذيل الأيمن والأخرى في الذيل الأيسر (طرفين) للتوزيع العيني تحت الفرض العدمي هذه القيم تحجز مساحة في كل ذيل مقدارها $\alpha/2$ ،إذا وقعت القيمة المحسوبة لإحصائي الاختبار من بيانات العينة بين هاتين القيمتين نقبل الفرض العدمي إما خلاف ذلك نقبل الفرض البديل .
- إذا كان الاختبار من طرف واحد (طرف ايمن) فان منطقة الرفض تكون من جهة اليمين فنحسب القيمة الجدولية التي تحجز مساحة في الذيل الأيمن مقدارها α، إذا كانت القيمة المحسوبة اقل من القيمة الجدولية نقبل العدمي والعكس صحيح. إذا كانت منطقة الرفض جهة اليسار (طرف أيسر) نحسب القيمة الجدولية التي تحجز مساحة في الذيل الأيسر مقدارها α ، إذا كانت القيمة المحسوبة اكبر من القيمة الجدولية نقبل العدمي والعكس صحيح .

الطريقة الثانية :

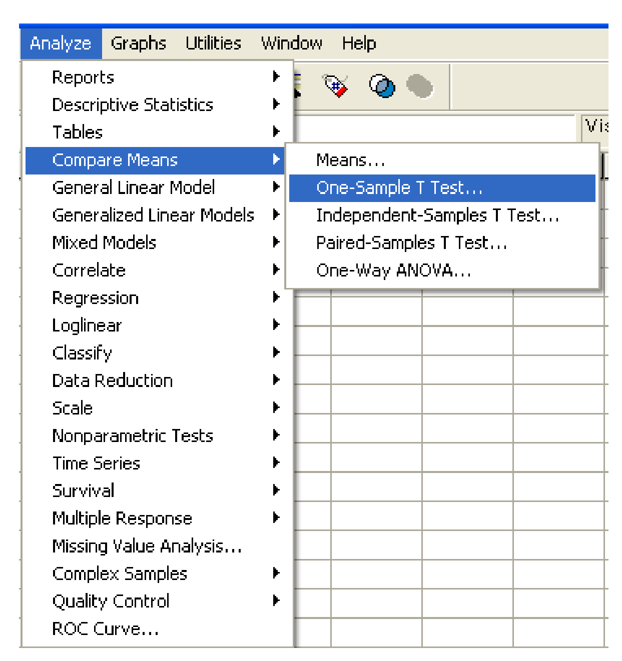
هناك طريقة أخرى للرفض أو للقبول استخدمت كثيرا في الآونة الأخيرة وتعتمد على احتمال محسوب يسمى*P-Value* ويرمز له في الحزمة SPSS بالرمز .Sig ، وتعرف

P-Value علـى أنها مستوى معنويـة محسوب أو خطأ من النـوع الأول محسوب، هذه الطريقة تستخدم كثيرا في البرامج الجاهزة، تتلخص هذه الطريقة في الآتي:

- يجـرى الاختبـار بـنفس الخطـوات السـابق الحـديث عنهـا ، لا تحسـب قـيم جدوليـة لإحصائي الاختبار ولكن نستخدم احتمال محسوب يرمز لـه بـالرمز .Sig، تعبر هذه القيمـة عـن مسـاحة معيـنة فـي ذيـل (أو ذيلـين) للتوزيـع العينـي لإحصـائي الاختبـار وتسمى قيمة مستوى المعنوية المحسوب .

إذا كان الاختبـار من طـرفين تقـارن هذه القيمـة بقيمـة $\alpha/2$ ، إذا زادت القيمـة المحسوبة عن($\alpha/2$) نقبل الفرض العدمي والعكس صـحيح. إذا كانـت الاختبـار من طـرف واحد نقارن القيمة المحسوبة بقيمة α ، إذا زادت عنها نقبل الفرض العدمي والعكس صحيح .

تظهر أوامر اختبارات معلميه بـالنقر علـى Analyze فـي شريط الأوامر كما فـي الشكل التالـي:



الشكل (١)

اختبار عينة واحدة

في اختبار عينة أو عينتين نحتاج إلى إحصائي للاختبار يطلق عليها اسم *t* نسبة إلـى التوزيـع الاحتمـالي الخـاص بـه والمسمـى باسم *t-Distribution* التوزيـع، *t* مـن

التوزيعات المتماثلة ويشبه كثيرا التوزيع الطبيعي، يؤول هذا التوزيع إلى التوزيع الطبيعي عندما يزيد حجم العينة (أكبر من ٣٠ مفردة) معنى ذلك أنه في الحالات التي يطبق فيها إحصائي الاختبار t يستبدل بالتوزيع الطبيعي عندما يزيد حجم العينة عن ٣٠ مفردة. اختبار t وهو اختبار معلمي يستخدم هذا الاختبار لفحص فرض يتعلق بالوسط الحسابي .

ويجب تحقق الشرطين التاليين حتى نستخدم اختبار t:

- يجب أن يتبع توزيع البيانات التوزيع الطبيعي، ويستعاض عن هذا الشرط بزيادة حجم العينة إلى أكثر من ٣٠ مفردة.
- يجب أن تكون العينة عشوائية أي لا تعتمد مفرداتها على بعضها.

اختبار عينة واحدة يتوقف على فرضين أحدهما يسمى الفرض العدمي والآخر يسمى الفرض البديل، تدور هذه الفروض غالبا حول متوسط المجتمع أو تباين المجتمع أو النسبة في المجتمع ...الخ .

ولإجراء اختبار متوسط عينة نفرض أن متوسط المجتمع (المعلمة) مجهول و نريد اختبار بعض الفروض التي تدور حول المعلمة المجهولة، تحديد إحصائي الاختبار سيتوقف على هل تباين المجتمع المسحوبة منه العينة معروفا (فرض نظري لا يمكن تحقيقة) أم غير معروف. إذا كان تباين المجتمع معروفا فإن التوزيع العيني الذي يتحكم في عمليات الرفض والقبول هو التوزيع الطبيعي، إذا لم يكن تباين المجتمع معروف يكون البديل الوحيد للتوزيع الطبيعي هو توزيع t. و هو اختبار خاص بمقياس الفترات إذا كانت العينة واحدة وصغيرة أي أن حجم العينة أقل من ثلاثين، ونحاول هنا معرفة هل يختلف المتوسط الذي تم احتسابه من خلال العينة بقيمة ذات دلالة إحصائية عن متوسط المجتمع الذي أخذت منه العينة.

- الفروض الإحصائية لاختبار t:

فرض العدم : $H_0 : \mu = \hat{\mu}$

الفرض البديل: $H_A : \mu \neq \hat{\mu}$

حيث: (μ) هي متوسط المجتمع وتقرأ "ميو"

($\hat{\mu}$) هي قيمه متوسط العينه وتقرأ "ميو هات"

- **شروط استخدام توزيع t:**

- يجب أن يتبع توزيع البيانات التوزيع الطبيعي، ويعوض عن هذا الشرط بزيادة حجم العينة إلى أكثر من ٣٠ مشاهده.

- يجب أن تكون العينة عشوائية أي لا تعتمد مشاهداتها على بعضها.

- إحصاءه t المستخدمة في الاختبار:

$$t = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

حيث: $\overline{X}$ متوسط العينه وتقرأ "اكس بار"

S الانحراف المعياري للعينة،

n حجم العينة.

- **الفرق بين اختبار t واختبار Z:**

في حالة العينات الصغيرة (n<30) وتباين المجتمع مجهول يتم استخدام اختبار t وفي حالة العينات الكبيرة ($n \geq 30$) يتم استخدام اختبار Z. أي أن توزيع t يؤول إلى توزيع Z عندما يزيد حجم العينة عن ٣٠ مشاهدة.

مثال (١):

قررت لجنة فحص طلبية من أورق الطباعة من النوع ٨٠ جرام التأكد من أن المتوسط الحسابي لعدد الأوراق في الرزمة الواحدة لا يساوي ٥٠٠ ورقة، أخذت عينة عشوائية من الطلبية من عشرة رزم (١٠ رزم) وتم وزن كل زرمة فكانت النتائج كالأتي:

جدول (١)

الرزمة	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠
الوزن بالكليو جرام	٣.٩	٤.١	٣.٨	٣.٧	٤	٤.٢	٣.٩	٣.٩	٤	٣.٧

المطلوب ما يلي :

- هل ترفض اللجنة الطلبية بناء على هذه المعلومات عند مستوى معنوية ٠.٠٥ و٠.٠١ .

⌖ تقدير فترة ثقة ٩٩% للوسط الحسابي للمجتمع .

الحل :

إذا كانت الرزمة تحتوي فعلا على ٥٠٠ ورقة ووزن الورقة الواحدة ٨٠ جرام فإن وزن الرزمة يجب أن يساوي ٤ كجم (٥٠٠×٨٠ جرام = ٤٠٠٠ جرام) وبفرض أن وزن الرزمة يتوزع كالتوزيع الطبيعي فان قرار اللجنة يجب أن يكون مبنياً على الاختبار الإحصائي التالي:

⌖ الفروض الإحصائية:

H_0 : متوسط وزن الرزمة = المتوسط المفترض ٤ .

H_A : متوسط وزن الرزمة لا يساوي المتوسط المفترض (الاختبار من طرفين).

$$H_0 : \mu = 4\,km$$

$$\mathrm{H_A} : \mu \neq 4 \quad km$$

⌖ احصاءة الاختبار هي:

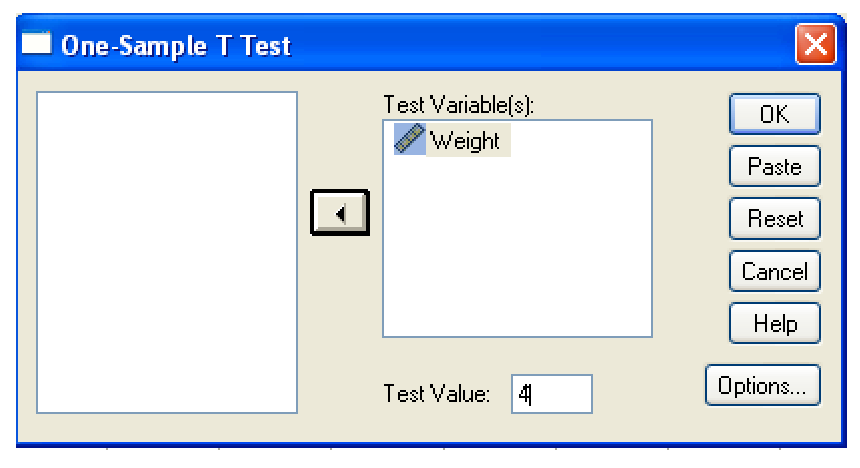
$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

(لأن تباين المجتمع مجهول وحجم العينه أقل من ٣٠)

حيث $\overline{X}$ يمثل الوسط الحسابي للعينة وS يمثل الانحراف المعياري للعينة، n حجم العينة و μ الوسط الحسابي للمجتمع بموجب فرضية العدم ويساوي ٤ وهذه الإحصائية تتبع توزيع t بدرجة حرية $\upsilon = n-1$، تحتسب الإحصائية $\overline{X}$ وهي دالة الاختبار وكذالك s من قبل البرنامج) :

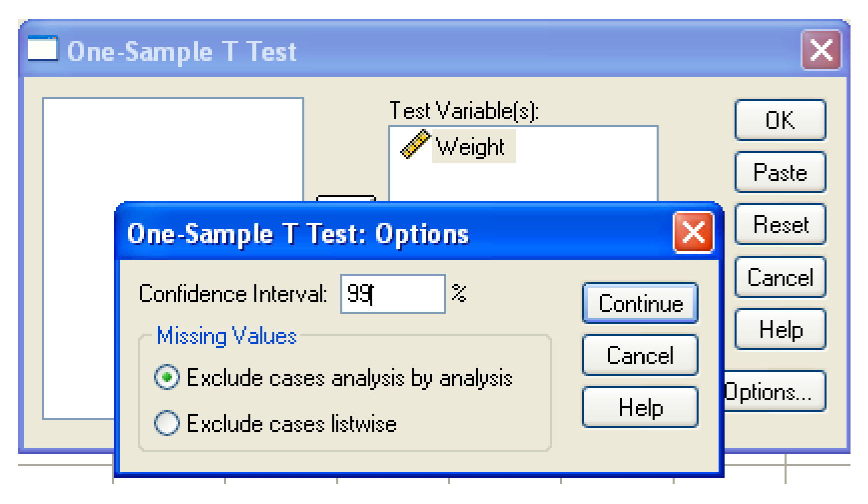
لتنفيذ الاختبار نتبع الخطوات التالية :

← من شريط القوائم نختار Analyze ومن قائمة الأوامر نختار Compare Means ونختار الأمر الفرعي One Sample t-test فيظهر صندوق حوار One Sample T-test انظر الشكل (٢):



الشكل(٢)

← انقر Options فيظهر صندوق حوار Options الذي نرتبه أنظر الشكل (٣):



الشكل (٣)

يمكن بواسطة Options تحديد فترة الثقة (المطلوب الثاني في السؤال) حيث نقوم بكتابة الفترة ٩٩% في حقل Confidence Interval وذلك لحساب ١% مستوى معنوية فيترتب عليه درجة ثقة مقدارها المكمل لهذا الرقم أي انه ٩٩%.

أما Missing Values فيستفاد منه في التعامل مع القيم المفقودة.

← ننقر Ok في صندوق حوار One sample T-test فتظهر النتائج التالية :

جدول (٢)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Weight	10	3.9200	.16193	.05121

جدول (٣)

One-Sample Test

	Test Value = 4					
	t	df	Sig. (2-tailed)	Mean Difference	99% Confidence Interval of the Difference Lower	99% Confidence Interval of the Difference Upper
Weight	-1.562	9	.153	-.08000	-.2464	.0864

⌖ **الطريق الأولى:**

حيث قيمه احصاءة t المحسوبة تساوي t=-1.5622 وهذه تعرف بـ t المحسوبة [$t_{calc.}$] ويجري مقارنتها بقيمة t الجدولية [$t_{tab.}$] بدرجة حرية $\nu = 9$ ولمستوى معنوية $\alpha/2$ (لأن الاختبار من طرفين) هي كالأتى

$$T = t_{(9,0.025)} = 2.262 \qquad (\alpha = 0.05)$$

$$T = t_{(9,0.005)} = 3.250 \qquad (\alpha = 0.01)$$

وحيث أن القيمة المطلقة (بدون اشارة) لـ t المحسوبة (1.5622) تقع بين حدود القيمة الجدولية ناحية اليسار واليمين ،إذن نقبل فرضية العدم ونأخذ الفرضية البديلة أي أن الوسط الحسابي للمجتمع يختلف معنويا عن القيمة 4 لمستوى معنوية 5% و1% .

⌖ **الطريقة الثانية:**

للاختبار تعتمد على قيمة p-value وتتميز عن الطريقة الأولى كونها لا تحتاج إلى استخدام جداول التوزيعات ويتم احتسابها مباشرة من قبل SPSS وهي القيمة .sig

في الجدول أعلاه ويمكن تعريف p-value بأنها أقل قيمة لـ α التي يتم عندها رفض فرضية العدم، حيث نرفض فرضية العدم عندما .sig أقل من α .

وبما أن sig.>0.05 و sig>0.01 إذن نقبل الفرض العدم لمستوى معنوية ٠.٠٥ و ٠.٠١.

أما بالنسبة لفترة ثقة ٩٩% للفرق بين متوسط العينة ومتوسط المجتمع فيمكن كتابتها كما يلي :

$$(-0.2464 \le \overline{X} - \mu \le 0.0864) = 99\%$$

أما تقدير فترة ثقة ٩٩% لمتوسط المجتمع μ كما يلي :

$$(\overline{X} - t_{(9,0.005)}(s/\sqrt{n}) \le \mu \le \overline{X} + t_{(90.005)}(s/\sqrt{n})) = 99\%$$

$$P(3.753578 \le \mu \le 4.086422) = 99\%$$

أي أن احتمال أن يقع متوسط المجتمع بين القيمتين 3.753578 و 4.086422 يساوي 99% .

وعليه نقبل الفرض العدم بمستوى معنوية ٠.٠١ لأن القيمة 4 تقع داخل فترة الثقة وهذه تعتبر طريقة ثانية لاختبار الفرضية.

اختبار عينتين مستقلتين Independent Grouped t-Test:

نفرض أن لدينا عينتين مستقلتين وكنا مهتمين بمتغير معين في العينتين وكنا نريد اختبار هل متوسط المجتمع المسحوب منه العينة الأولى يساوي متوسط المجتمع المسحوب منه العينة الثانية.

- الفروض الإحصائية:

فرض العدم $H_0 : \mu_1 = \mu_2$

الفرض البديل $H_1 : \mu_1 \ne \mu_2$

- عند اختيار Independent-Sample T Test يجب مراعاة الشروط التالية:
  - ✓ العينتين مستقلتين .
  - ✓ البيانات في كل عينة تتبع التوزيع الطبيعي.
  - ✓ العشوائية في اختيار العيينتين.

✓ المجتمعين المسحوبة منهما العينتين متجانس بمعنى أن تباين المجتمع المسحوب منه العينة الأولى يساوي تباين المجتمع المسحوب منه العينة الثانية ، وللتأكد من التجانس يجب إجراء اختبار سابق لاختبار T يسمى اختبار التجانس ويكون الفرض العدمي والبديل للاختبار كالتالي :

← الفرض العدمي هناك تجانس (تباين المجتمع الأول يساوي تباين المجتمع الثاني).

← الفرض البديل لا يوجد تجانس (تباين المجتمع الأول لا يساوي تبين المجتمع الثاني). والذي ممكن أن يكونا على الشكل التالي:

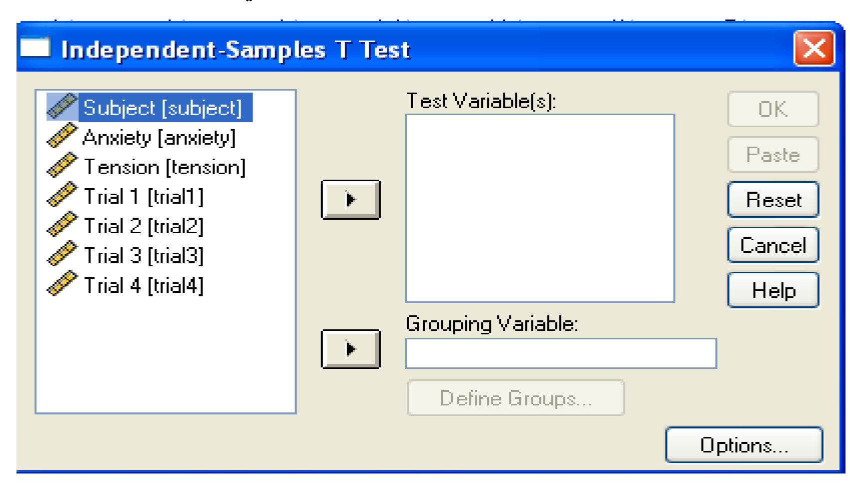
$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

أ- إذا تم قبول فرض العدم أي أن هناك تجانس سنستمر في إجراء اختبار t.

ب- إذا تم قبول الفرض البديل أي عدم وجود التجانس فإنه لا يجوز استخدام اختبار ويستبدل باختبار آخر شبيه باختبار t.

ج- اختبار التجانس يتم بواسطة اختبار آخر يسمى F-test، البرنامج يجري اختبار التجانس ويعطي نتائج اختبار t إذا كان هناك تجانس ويعطي أيضا نتائج لاختبار الشبيه باختبار t إذا لم يكن هناك تجانس وعلى المستخدم اختيار النتيجة الملائمة له، وعند تفعيل هذا الخيار يظهر الشكل التالي:



الشكل (٤)

المثال التالي يوضح كيفية تطبيق هذا الاختبار في الواقع العملي:

مثال(٢):

أخذت عينتين عشوائيتين من مجموعة متشابهة من الأطفال وأعطي أطفال العينة الأولى غذاء رمزنا لها X و أعطي أطفال العينة الثانية غذاء آخر رمزنا لها Y وكانت الزيادة في أوزان الأطفال بالكيلو جرام في العينتين بعد مدة معينة كما في الجدول التالي:

جدول (٤)

العينة الأولي	٢.٥	١.٥	٥.٥	٤.٥	٣.٥	-
العينة الثانية	٢	١.٥	٠.٥	١.٥	١	٢.٥

اختبر فرض عدم وجود فرق بين أثر الغذائيين " X " و " Y " في متوسط زيادة وزن الأطفال عند مستوي معنوية ٠.٠٥.

الحل :

يتضح من المعطيات أن عدد الأطفال في العينة الأولى ٥ وعدد الأطفال في العينة الثانية ٦ وليس من الضروري أن تتساوى أحجام المجموعتين .

لإجراء اختبار الفرق بين متوسطي عينتين نتبع الخطوات التالية:

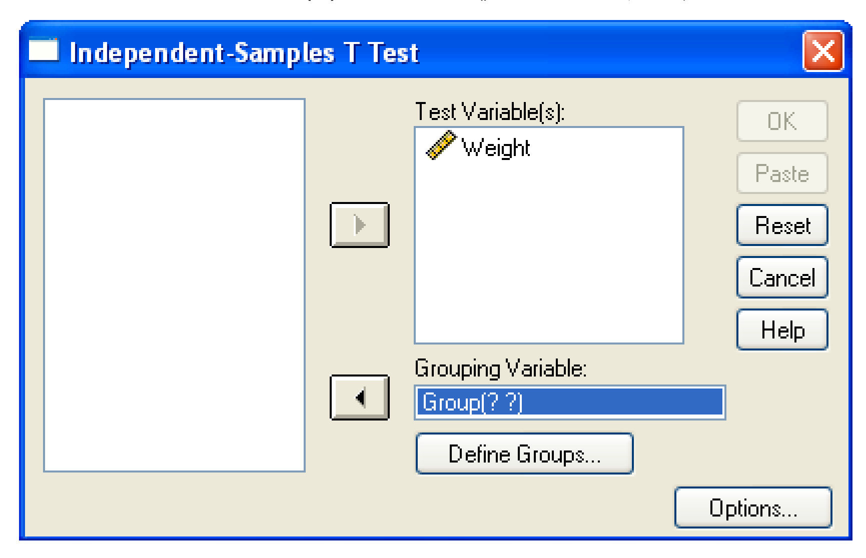
← يجب إدخال البيانات بطريقة معينة في شاشة المحرر وهي الطريقة التي يتعامل بها البرنامج كالتالي :

Weight	Group
2.50	X
1.50	X
5.50	X
4.50	X
3.50	X
2.00	Y
1.50	Y
.50	Y
1.50	Y
1.00	Y
2.50	Y

الشكل (٥)

العمود الأول يحتوي على جميع الدرجات للطلاب والطالبات و العمود الثاني يحتوي على A أي أن الدرجة المقابلة له من المجموعة الأولى (طلاب) و B أي أن الدرجة المقابلة له من المجموعة الثانية (طالبات) .

← من القائمة Analyze ننقر الأمر Compare Means ثم ننقر الأمر الفرعي Independent sample test ليفتح صندوق حوار Independent sample T-test الذي نقوم بترتيبه كالتالي انظر الشكل (٦) :



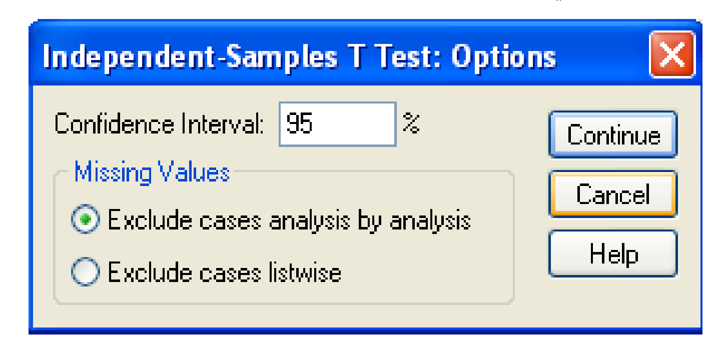
الشكل (٦)

قمنا بنقل المتغير Weight ونقلناه إلى variable test و المتغير Group (متغير التجزئة) في قائمة Grouping variables ، ثم تعريف المجاميع X و Y عن طريق الزر Define Groups كالتالي انظر الشكل (٧):

Define Groups
Group 1: X
Group 2: Y
Continue
Cancel
Help

الشكل (٧)

← يمكن بواسطة النقر على Options تحديد فترة الثقة حيث نقوم بكتابة الفترة المطلوبة في حقل Confidence Interval وهنا فترة ثقة مقدرها ٩٥% عن طريق الشكل الآتي:



الشكل (٨)

أما Missing values فيستفاد منه في التعامل مع القيم المفقودة فتوجد طريقتين لتقدير يمكننا الاختيار بينهما.

← انقر على Continue للعودة إلى الصندوق الأصلي .

← انقر على الأمر Ok للتنفيذ .

تم الحصول على النتائج التالية :

أ- الجدول الآتي يوضح حجم المجموعة والوسط و الانحراف المعياري والخطأ المعياري لكل مجموعة .

جدول (٥)

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Weight	X	5	3.5000	1.58114	.70711
	Y	6	1.5000	.70711	.28868

ب- الجدول التالي به اختبارين الأول اختبار التجانس والثاني اختبار T وهو كالآتي :

جدول (٦)

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Weight	Equal variances assumed	3.165	.109	2.803	9	.021	2.00000	.71362	.38567	3.61433
	Equal variances not assumed			2.619	5.326	.044	2.00000	.76376	.07229	3.92771

ويوضح الجدول من اليسار إلى اليمين الآتى :

- العمود الأول يسارا به اسم المتغير Weight و العمود الثاني والثالث يسار لإجراء اختبار التجانس كالآتي :

- فرض العدم : هناك تجانس Equal Variances.
- الفرض البديل: هناك عدم تجانس Equal Variances not assumed.
- قيمة مستوى المعنوية المحسوبة Sig.=0.109 وهي تزيد على مستوى المعنوية المحدد وهو٠.٠٥ اذن نقبل الفرض العدمي القائل بأن هناك تجانس ونرفض الفرض البديل.
- العمود الرابع والخامس لإجراء إما اختبار *t* أو الاختبار الشبية باختبار t، وحيث أننا قبلنا أن هناك تجانس نتعامل مع اختبار t والنتائج في الصف الأول من الجدول ونهمل الثاني لأنه خاص بالاختبار الشبية باختبار t .
- من الصف الأول نجد نتائج اختبار t ،قيمة مستوى المعنوية المحسوبة هي Sig.=0.021 وهي أقل من ٠.٠٥ إذن نرفض الفرض العدمي القائل أنه يوجد فرق بين مستوى الطلاب والطالبات في المجموعتين.
- يوجد أيضا حدود ثقة للفرق بين متوسطي المجتمعين يمكن للباحث استخدامها إذا رغب وكذلك خطأ الفرق .

اختبار عينتين غير مستقلتين Paired Sample Test:

نتناول هنا موضوع اختبارات الفروض عن فرق بين متوسطي مجتمعين عندما يكون العينتين المسحوبتين منهما غير مستقلتين ويكون مشاهدات العينتين عبارة عن مشاهدات أزواج مرتبة من القيم $[(X_r, Y_r), r = 1,2,..., N]$ حيث X_1 تمثل قيمة المشاهدة للعينة الأولى مأخوذة من المفردة رقم ١ و Y_1 تمثل قيمة المشاهدة للعينة الثانية المأخوذة من المفردة رقم ١ وبصفة عامة فان X_r تمثل قيمة المشاهدة العينة الأولى المأخوذة عن المفردة رقم r بينما Y_r تمثل قيمة المشاهدة للعينة الثانية المأخوذة عن المفردة رقم r أي أنه N من المفردات وكل مفردة مسجل لها مشاهدتين (أو قيمتين أو قياسيين) أولهما للعينة الأولى و ثانيهما للعينة الثانية. مثل تسجيل قياسات ضغط الدم لكل شخص في عينة مكونة من ١٠ أشخاص مرة قبل تعاطي دواء معين ومرة أخرى بعد ذلك الدواء بعد أسبوعين أو ما يطلق عليه في الدراسات التجريبية الاختبار القبلي (Pre test) والاختبار البعدي (Post test) على نفس العينة. وبصفة عامة إذا كان لدينا N من أزواج المشاهدات المرتبة $(X_r, Y_r), r = 1,2,..., N$ مأخوذة على N من المفردات فإنه يمكن اعتبار أن: $X_1, X_2, ..., X_N$ عينة مسحوبة من مجتمع متوسطه μ_1 وان $Y_1, Y_2, ..., Y_N$ عينة مسحوبة من مجتمع متوسطه μ_2 .

يستعمل هذا الاختبار لاكتشاف معنوية (دلالة) الفرق بين متوسطي متغيرين غير مستقلين لمجموعة (عينة) واحدة، معنى ذلك أن لدينا عينة واحدة وكل مفردة في العينة تعطى قراءتين القراءة الأولى تمثل العينة الأولى والقراءة الثانية تمثل العينة الثانية.

يعتمد الاختبار على إيجاد الفرق بين القراءات المتناظرة في العينتين (نرمز إلى مجموع الفرق بين العينتين بالرمز d).

ويكون صياغة الفروض كالتالي :

فرض العدم $H_0 : d = 0$

الفرض البديل $H_A : d \neq 0$

- **شروط الاختبارات للعينتين المترابطة:**

١. الشروط الأساسية لاختبار t.

٢. الفرق d بين العينتين يتبع التوزيع الطبيعي وذلك إذا كان $(n<30)$.

أما إذا كان $(n\geq 30)$ فهذا الشرط يسقط.

المثال يوضح كيفية تطبيق هذا الاختبار في الواقع العملي:

مثال(٣) :

عقدت دورة تدريبية لمجموعة من الموظفين ولمعرفة مدى استفادة المتدربين من الدورة أجري لهم اختبارين الأول (القبلي) في بداية الدورة والثاني (البعدي) في نهايتها. وفيما يلي بيان بنتيجة الاختبارين لعينة عشوائية مكونه من تسعة من المتدربين:

جدول (٦)

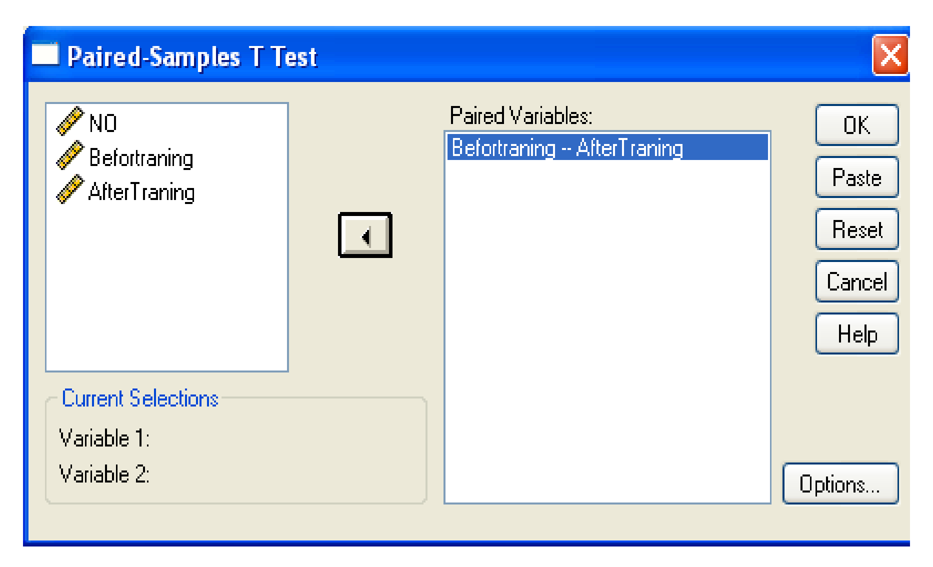
رقم المتدرب	١	٢	٣	٤	٥	٦	٧	٨	٩
الدرجة قبل التدريب	٧	٩	٤	٧	٦	٨	٦	١١	٦
الدرجة بعد التدريب	١٥	١٨	١٦	١٨	١٤	١٦	١٨	٢٠	١٩

اختبر فرض أن الدورة التدريبية ليس لها أثر في رفع مستوى المتدربين عند مستوى معنوية ٥ % .

الحل :

لتنفيذ الاختبار نتبع الخطوات التالية :

← من القائمة Analyze انقر الأمر Compare Means ثم انقر الأمر الفرعي Paired – Sample T test فيظهر صندوق حوار Paired – Sample T test الذي نقوم بترتيبه انظر الشكل (٩) :



الشكل (٩)

- ✓ المتغيرات قبل التدريب وبعد التدريب يمثلان مشاهدات النوعين قبل التدريب وبعد التدريب حيث يتوجب إدخال كلا المتغيرين في نفس الوقت في قائمة *Paired variables*.
- ✓ يمكن بواسطة النقر على *Options* تحديد فترة الثقة حيث نقوم بكتابة الفترة المطلوبة في حقل *Confidence Interval*.
- ✓ أما *Missing values* فيستفاد منه في التعامل مع القيم المفقودة.
- ← انقر على *Continue* للعودة إلى الصندوق الأصلي .
- ← انقر على الأمر *Ok* للتنفيذ .

تم الحصول على النتائج التالية :

- ⌖ الجدول التالي يحتوي على بيانات إحصائية عن المتغيرين وتتضمن الوسط وعدد الحالات و الانحراف المعياري و الانحراف المعياري للوسط .

جدول (٧)

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Befortraning	7.1111	9	2.02759	.67586
	AfterTraning	17.1111	9	1.96497	.65499

⌖ الجدول التالي يحسب معامل الارتباط بين البيانات قبل وبعد ويعطي اختبار لمعامل الارتباط حيث يكون الفرض العدمي الارتباط يساوي صفر والفرض البديل الارتباط لا يساوي صفر .

معامل الارتباط موجب وهو ٠.٤٩٩ أما قيمة $Sig.=0.172$ وهي أكبر من ٠.٠٥ نقبل فرض العدم القائل بأن الارتباط غير معنوي.

جدول (٨)

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Befortraning & AfterTranin	9	.499	.172

جدول (٩)

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	Befortraning - AfterTraning	-10.0000	2.00000	.66667	-11.5373	-8.46266	-15.000	8	.000

الجدول التـالي يعطـي اختبـار t للفرق بين عينتين غيـر مستقلتين ، نلاحظ أن Sig.=.000 آخر عمود وهـي أقل من ٠.٠٥ إذن نـرفض الفرض العدمي القائل أنـه لا يوجد فرق بين العينتين ونقبل الفرض البديل القائل أن هنـاك فرق درجـات حدث نتيجـة التدريب. هنـاك بيانات أخرى بالجدول يمكن للباحث الاستفادة منها إذا رغب.

الفصل الخامس

تحليل التباين

Analyzes of Variance (ANOVA)

في دراستنا لاختبارات الفروض في الفصل السابق، كانت الدارسة قاصرة على مجتمع واحد أو مجتمعين على الأكثر. وشملت تلك الدارسة اختبار فروض عن متوسط مجتمع واحد واختبارات فروض بخصوص تساوي متوسطي مجتمعين. ومن الطبيعي تعميم هذا الأسلوب إلى حالة وجود m من المجتمعات ($m \geq ٢$) ويكون المطلوب منا مقارنة متوسطاتها. فمثلا إذا كان المطلوب إجراء مقارنة بين متوسطات إنتاجية الفدان لخمسة أصناف مختلفة من القطن، رمزنا لمتوسطات إنتاج الفدان من هذا الأصناف الخمسة بالرموز: $\mu_1 = \mu_2 = \ldots = \mu_5$ يكون فرض عدم وجود فرق بين هذه المتوسطات (أي فرض تساوي متوسطات هذا الأصناف الخمسة) هو:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \qquad (١)$$

ويكون الفرض البديل H_A هو أن فرض العدم ليس صحيحا أي أن H_A هو فرض عدم تساوي متوسطين على الأقل من هذه المتوسطات ويكون:

$\mathrm{H_A}$: يوجد اثنين على الأقل من المتوسطات $\mu_1, \mu_2, \ldots, \mu_5$ غير متساويين (٢)

ويستخدم أسلوب تحليل التباين في اختبار فرض تساوي متوسطات عدة مجتمعات ضد فرض عدم تساويها فمثلا يكون فرض العدم والفرض البديل كما في (١)، (٢) وتتضح مدى أهمية أسلوب تحليل التباين في اختبار فرض عدم مثل H_0 في (١) ضد فرض بديل مثل H_A بملاحظة أنه لإجراء مثل هذا الاختبار بإتباع الأسلوب السابق عرضه (باستخدام اختبار t) لمقارنة متوسطي كل مجموعتين على حدة فإن هذا سوف يستلزم إجراء عشرة اختبارات من ذلك النوع ولاشك أن هذا سوف يحتاج إلى كثير من الوقت والجهد. ومن الواضح أنه كلما زاد عدد المجتمعات المطلوب مقارنة متوسطاتها كلما زادت صعوبة إجراء المقارنات المطلوبة بإتباع أسلوب مقارنة كل متوسطين على حدة. فمثلا إذا كان مطلوب مقارنة متوسطات عشرة مجتمعات فإن ذلك يستلزم إجراء ٤٥ اختبار كل منها لمقارنة متوسطين فقط باستخدام اختبار t. ويستخدم أسلوب تحليل التباين لمقارنة متوسطات عدة مجتمعات في العديد من البحوث مثل مقارنة متوسطات إنتاج عدة أصناف من محصول معين، مقارنة متوسطات عدة الأيام اللازمة للشفاء من مرض معين لعدة أنواع من الأدوية ، مقارنة عدة طرق من طرق التدريس ، مقارنة عدة طرق لإنتاج سلعة معينة ، مقارنة متوسط الدخل الشهري للأسرة بعدة محافظات وهكذا.

وفي تحليل التباين يستخدم تعبير معالجات Treatments ليعبر عن التصنيفات المختلفة المستخدمة والمطلوب مقارنة متوسطاتها فقد تكون المعالجات أنواع مختلفة من الأدوية أو قد تكون أنواع من الأسمدة أو طرق مختلفة للتدريب أو محافظات مختلفة ومطلوب إجراء مقارنات بين متوسطات ظاهرة ما باستخدام عينة عشوائية من كل منها.

وباختصار يقصد بتحليل التباين العمليات الرياضية الخاصة بتقسيم مجموع المربعات الكلي لمجموعة من البيانات إلى مصادره المختلفة وتلخص نتائج التحليل في جدول يعرف بجدول تحليل التباين ANOVA Table. إن الهدف من إجراء هذا التحليل هو اختبار فرضية تساوي متوسطات مجموعة من العينات وتعرف بالمعالجات أو المعاملات Treatment دفعة واحدة ولهذا فهو يعتبر توسيعاً لاختبار t.

- **اختبار ف:Test-F**

اختبار تحليل التباين يعتمد على إحصائي اختبار يطلق عليه اسم توزيع F نسبة إلى التوزيع الاحتمالي F Distribution.

ولاختبار مساواة متوسطات المجموعات يتم تقسيم التباين الكلي للمتغير التابع إلى مركبتين الأولى معروفة المصدر وتسمى بين المجموعات (Between Group) ومصدرها الفرو ق بين متوسطات المجموعات، فإذا كان هذا الجزء كبيرا فإن متوسطات المجموعات غير متساوية، والثانية داخل المجموعات (Within Group) وهي الجزء غير معروف المصدر والذي يسمى في بعض الأحيان الباقي Residuals أو الخطأ Error.

- ✓ متى نرفض الفرضية التي تقول: أن متوسطات المجموعات متساوية؟

نرفض هذه الفرضية إذا كانت نسبة التباين بين المجموعات (معروف المصدر) إلى التباين داخل المجموعات (غير معروف المصدر) كبيرا، وهذه النسبة تسمى قيمة F، فإذا كانت قيمة F كبيرة نسبيا فإن متوسطات المتغير التابع للمجموعات غير متساوية، ولكن إلى أي حد تعتبر قيمة F كبيرة حتى نرفض الفرضية التي تقول أن متوسطات المجموعات متساوية؟

- ✓ متى نقول أن قيمة F كبيرة نسبيا؟

إذا كانت المساحة فوقها (مستوى الدلاله Sig) أقل من المستوى المقبول لدينا α والتي غالبا تساوي ٠.٠٥ فإذا كانت قيمة Sig أقل من α فإن متوسطات المجموعات غير متساوية، وإذا كانت قيمة .Sig أكبر من $\alpha = 0.05$ فإن متوسطات المجموعات متساوية.

تحليل التباين في اتجاه واحد (الأحادي) (One-Way ANOVA):

يقصد بتحليل التباين العمليات الرياضية الخاصة بتقسيم مجموع المربعات الكلي لمجموعة من البيانات إلى مصادره المختلفة وتلخص نتائج التحليل في جدول يعرف بجدول تحليل التباين ANOVA Table . إن الهدف من إجراء هذا التحليل هو اختبار فرضية تساوي متوسطات مجموعة من العينات وتعرف بالمعالجات أو المعاملات Treatment دفعة واحدة.

✓ **متى يستخدم تحليل التباين في اتجاه واحد؟**

يستخدم إذا كان لدينا عددا من العينات ونريد أن نختبر هل هذه العينات مسحوبة من مجتمعات متوسطاتها متساوية أم لا؟ بمعنى آخر هل هذه العينات مسحوبة من نفس المجتمع أم لا؟ يمكن اختبار ذلك الفرض باستخدام اختبار تحليل التباين، أي أن:

⌖ **الفروض الإحصائية:**

الفرض الصفري

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

ضد الفرض البديل

H_A : يوجد متوسطان على الأقل غير متساوين

يمكن تلخيص خطوات الاختبار كالتالي:

⌖ تحسب مجموع مشاهدات كل عينة وكذلك متوسطها ، نفرض أن $\sum_{i=1}^{N} X_{ir}$ ترمز لمجموع مشاهدات العينة r ، $\overline{X}_r = \left(\sum_{i=1}^{N} X_{ir}\right) / N$ ترمز إلى وسطها الحسابي، حيث أن $r = 1,\ldots,k$ وعدد البيانات هي N فيمكن تنظيم حساب مجموع مشاهدات كل عينة ومتوسطها باستخدام جدول البيانات الآتي:

جدول (٤)

المشاهدات	المجتمع				
	1	2	3	...	k
1	X_{11}	X_{12}	X_{13}	...	X_{1k}
2	X_{21}	X_{22}	X_{23}	...	X_{2k}
3	X_{31}	X_{32}	X_{33}	...	X_{3k}
: :	: :	: :	: :	... ...	... ...
N_j	$X_{N_1 1}$	$X_{N_2 2}$	$X_{N_3 3}$	...	$X_{N_k k}$
$\sum$	$\sum_{i=1}^{N_1} X_{i1}$	$\sum_{i=1}^{N_2} X_{i2}$	$\sum_{i=1}^{N_3} X_{i3}$	...	$\sum_{i=1}^{N_k} X_{ik}$
$\overline{X}$	$\overline{X_1}$	$\overline{X_2}$	$\overline{X_3}$	...	$\overline{X_k}$

- يحسب المجموع الكلي لمشاهدات جميع العينـات ونرمـز لـه بـالرمز $\sum$ والمتوسط العام $\overline{X}$ كآلاتي:

$$\sum = \sum_{i=1}^{N_1} X_{i1} + \sum_{i=1}^{N_2} X_{i2} + \sum_{i=1}^{N_3} X_{i3} + \ldots + \sum_{i=1}^{N_k} X_{iK}$$

وأيضا $\overline{X} = \frac{\sum}{N}$ حيث أن $N = N_1 + N_2 + \ldots + N_k$.

- يحسب مجموع المربعات الكلي Total sum of squares كـآلاتي والـذي نرمـز لـه بالرمز SST:

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{N_i} X^2{}_{ji} - \frac{\sum^2}{N}$$

$$SST = X_{11}{}^{2} + X_{21}^{2} + \ldots + X_{N_1 1}^{2} + X_{12}^{2} + X_{22}^{2} + \ldots + X_{N_2 2}^{2} + \ldots + X_{N_k k}^{2} - \frac{\sum{}^{2}}{N} \qquad (٣)$$

أي أن مجموع المربعات الكلي نحصل عليه بإيجاد مجموع مربعات مشاهدات العينات مطروحا منه خارج قسمة مربع المجموع الكلي لمشاهدات جميع العينات على N.

ويمكن حساب مجموع المربعات الكلي بطريقة أخرى كالأتي:

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(X_{ji} - \overline{X} \right)^2$$

أي أن مجموع المربعات الكلي هو مجموع مربعات انحرافات كل مشاهدة من مشاهدات جميع العينات المأخوذة عن المتوسط العام لجميع هذه المشاهدات $\overline{X}$.

⌖ يحسب مجموع مربعات بين المعالجات Sum of squares Between treatments

مجموع مربعات بين المعالجات والذي نرمز له بالرمز SSR :

$$SSR = \frac{\left(\sum_{i=1}^{N_1} X_{i1} \right)^2}{N_1} + \frac{\left(\sum_{i=1}^{N_2} X_{i2} \right)^2}{N_2} + \ldots + \frac{\left(\sum_{i=1}^{N_k} X_{ik} \right)^2}{N_k} - \frac{\left(\sum \right)^2}{N} \qquad (٤)$$

ويمكن حساب مجموع مربعات بين المعالجات بطريقة أخرى كالأتي:

$$SSR = \sum_{i=1}^{K} N_i \left(\overline{X}_i - \overline{X} \right)^2$$

$$SSR = N_1 \left(\overline{X}_1 - \overline{X} \right)^2 + N_2 \left(\overline{X}_2 - \overline{X} \right)^2 + \ldots + N_k \left(\overline{X}_k - \overline{X} \right)^2$$

⌖ يحسب مجموع مربعات الأخطاء (مربعات داخل المعالجات)

Sum of Square within treatments or sum of square of Error

يحسب بالفرق بين مجموع المربعات الكلي ومجموع المربعات بين المعالجات و الذي تم حسابهم في الخطوتين السابقتين والذي نرمز له بالرمز SSE.

ويمكن حسابهم بطريقة أخرى كما يلي:

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{N_i} \left(X_{ji} - \overline{X}_i \right)^2 \qquad (٥)$$

أي أن مجموع مربعات داخل المعالجات يعبر عن التشتت داخل العينات أي عن تباعد مشاهدات كل عينة عن وسطها الحسابي.

- يحسب متوسط مربعات بين المعالجات $\sigma^2_{Between}$ وذلك بقسمة مجموع مربعات بين المعالجات على (عدد المعالجات-١) أي علي k-1، ويسمي k-1 بدرجات حرية بين المعالجات أي أن:

$$\text{MSR}= (k\text{-}1)/\text{SSR}= \sigma^2_{Between} \qquad (٦)$$

- يحسب متوسط مربعات داخل المعالجات σ^2_{within} وذلك بقسمة مجموع مربعات داخل المعالجات على (عدد المشاهدات الكلية-عدد المعالجات) أي على N-K، ويسمي N-K بدرجات حرية داخل المعالجات أي أن:

$$\text{MSE} = (\text{N-K})/\text{SSE}= \sigma^2_{within} \qquad (٧)$$

- يحسب خارج قسمة متوسط مربعات بين المعالجات على متوسط مربعات داخل المعالجات و إذا رمزنا بذلك بالرمز F للنسبة الناتجة فإن:

$$F = \frac{\sigma^2_{Betwen}}{\sigma^2_{within}} \qquad (٨)$$

ويكون لها توزيع F بدرجات حرية k-1 ، N–k. وتعتبر F المحسوبة هي إحصاءة الاختبار.

- إذا كانت F المحسوبة أكبر من $F_{(\alpha,k-1,n-k)}$ يرفض الفرض العدم H_0 و فيما عدا ذلك يقبل فرض العدم H_0.

والجدول الآتي الذي يعرض ملخص الخطوات السابقة يسمي جدول تحليل التباين analysis of variance table ويكون له الصيغة العامة المبينة في الجدول التالي:

جدول (٢)

مصدر الاختلاف	مجموع المربعات SS	درجات الحرية d.f	متوسط المربعات MS	F
بين المعاملات	SSR	**K-1**	**MSR**	$F_{calc} = \dfrac{MSR}{MSE}$
الخطأ	SSE	$\sum_{i=1}^{K} N_i - k$	**MSE**	
المجموع	SST	$\sum_{i=1}^{K} N_i - 1$		

والآن نتناول كيف يمكننا تنفيذ هذا باستخدام البرنامج مما يعمل علي إتاحة استخدام ذلك مع توفير الوقت والجهد في حالة زيادة عدد بيانات التحليل الإحصائي وذلك من خلال أمثلة تطبيقية متنوعة:

مثال(١):

استخدمت ثلاث طرق تعليمية مختلفة في تدريس مادة الحساب لثلاث عينات متشابهة من الطلاب وكانت درجاتهم في الامتحان النهائي كما بالجدول التالي:

جدول (٣)

العينة				
العينة الأولى	٨	٧.٥	٧	٥.٥
العينة الثانية	٩	١٠	٨	٩
العينة الثالثة	٦	٥	٥.٥	٣.٥

اختبر ما إذا كان هناك فروق معنوية بين متوسطات درجات الطلبة في الطرق الثلاث عند مستوى معنوية ٠.٠٥.

الحل:

في البداية يتم إدخال البيانات كلها في البرنامج في عمود واحد مع إضافة عمود ثاني يشير إلى رقم العينة فيكون شاشة data editor وقد سمناه factor نضع فيه قيمة ١

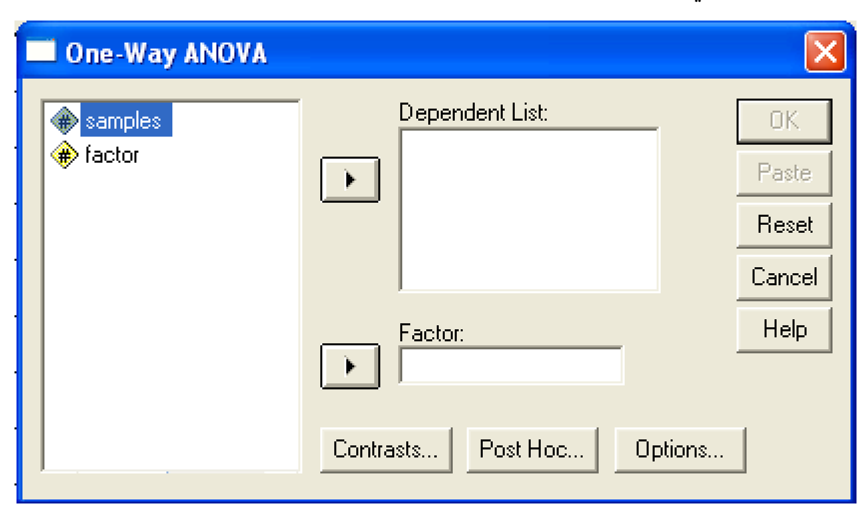
لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة الأولى ونضع القيمة ٢ لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة الثانية وهكذا ونضع القيمة ٣ لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة الثالثة وتكون البيانات كما في الشكل التالي:

samples	factor
8.00	1
7.50	1
7.00	1
5.50	1
9.00	2
10.00	2
8.00	2
9.00	2
6.00	3
5.00	3
5.50	3
3.50	3

وبعد إدخال البيانات نتبع الخطوات التالية:

- اختيار Compare means.
- اختيار One-Way ANOVA.

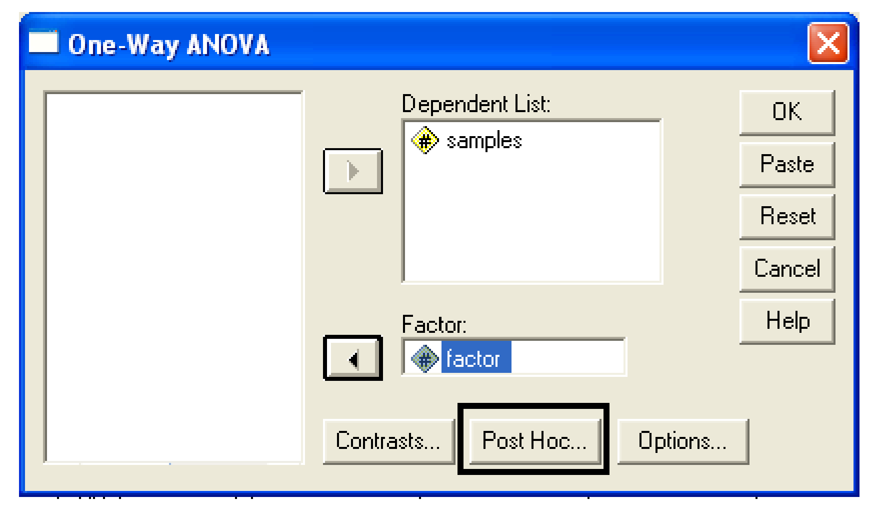
فيظهر الشكل الاتي:



شكل (١)

⌖ ننقل المتغيرات المراد حساب الفروق فيها إلى Dependent List ، ومتغير تصنيف المشاهدات في Factor.

فيظهر الشكل الآتي:



شكل (٢)

⌖ ننقر على OK تظهر النتائج كما في الجدول التالي:

جدول (٤)

ANOVA

SAMPLES

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	32.000	2	16.000	16.000	.001
Within Groups	9.000	9	1.000		
Total	41.000	11			

القرار الإحصائي:

يتضح من الجدول أن P-value = 0.001 وهي أقل من 0.05 عند F = 16.00 وهذا يدل على وجود فروق بين الطرق الثلاث لتدريس المهارات الحسابية.

نظراً لوجود فروق معنوية بين متوسطات الطرق فهذا يعني عدم تساوي متوسطي طريقتين على الأقل ولاختبار معنوية الفروق لكل زوج من المعالجات نلجأ إلى المقارنات المتعددة باستخدام طريقتي L.S.D و بإتباع الخطوات التالية:

باستخدام الخطوات السابقة حتى الشكل رقم (٢) فنقوم بنقر على post Hoc المحددة في الشكل (٢) فيظهر الشكل (٣):

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

LSD | S-N-K | Waller-Duncan
Bonferroni | Tukey | Type I/Type II Error Ratio: 100
Sidak | Tukey's-b | Dunnett
Scheffe | Duncan | Control Category: Last
R-E-G-W F | Hochberg's GT2 | Test
R-E-G-W Q | Gabriel | 2-sided < Control > Control

Equal Variances Not Assumed

Tamhane's T2 | Dunnett's T3 | Games-Howell | Dunnett's C

Significance level: .05

Continue | Cancel | Help

شكل (٣)

و نلاحظ وجود طرق عديدة للمقارنات المتعددة عندما يكون التباين لكل عينة متساوي حيث قمنا بتأشير LSD، ويوجد طرق للمقارنة عندما يكون التباين غير متساوي، وأخيراً تحديد مستوى المعنوية المطلوب 0.05 في Significance Level، عند نقر زر Continue كما نلاحظ أنه يوجد زر في الشاشة الرئيسة شاشة (٢) يطلق عليها Options و عند تفعيل هذا الاختيار نتمكن من حساب الإحصائيات الوصفية للقيم الفعلية أو القيم المفقودة كما في الشكل التالي:

One-Way ANOVA: Options

Statistics
Descriptive
Homogeneity-of-variance

Means plot

Missing Values
Exclude cases analysis by analysis
Exclude cases listwise

Continue | Cancel | Help

شكل (٤)

ثم زر Ok نحصل على المخرجات التالية:

جدول (٥)

Multiple Comparisons

Dependent Variable: SAMPLES
LSD

(I) FACTOR	(J) FACTOR	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interva	
					Lower Bound	Upper Bound
1	2	-2.0000*	.707	.020	-3.5996	-.4004
	3	2.0000*	.707	.020	.4004	3.5996
2	1	2.0000*	.707	.020	.4004	3.5996
	3	4.0000*	.707	.000	2.4004	5.5996
3	1	-2.0000*	.707	.020	-3.5996	-.4004
	2	-4.0000*	.707	.000	-5.5996	-2.4004

*. The mean difference is significant at the .05 level.

Post Hoc Tests

نلاحظ أنه باستعمال طريقة LSD فقد ظهرت فروق معنوية بمستوى معنوية 0.05 بين متوسطات المعالجات (١و٢) – (١و٣) و (٢و٣) حيث كانت قيمة p-value أو .sig أقل من 0.05 مع وجود علامة (*) على فروق المتوسطات ووجود تلك العلامة علامة وجود فروق بين العينات.

مثال (٢):

لمقارنة أربعة أنواع من الأدوية a،b ،c ،d لعلاج مرض معين اختيرت أربع عينات عشوائية كل منها مكون من خمسة أشخاص من المصابين بذلك المرض وأعطيت كل عينة نوع من الأدوية وكان عدد أيام العلاج اللازمة حتى الشفاء للأشخاص في العينات الأربعة كما بالجدول التالي:

جدول (٦)

العينة الأولى (دواء a):	٧	٣	٤	٦	٥
العينة الثانية (دواء b):	٦	٥	٩	٨	٧
العينة الثالثة (دواء c):	٣	٢	٥	٦	٤
العينة الرابعة (دواء d):	٨	٦	٣	٧	٦

اختبر فرض عدم وجود فرق معنوي بين متوسطات عدد الأيام اللازمة حتى الشفاء للأدوية الأربعة a، b، c، d وذلك عند مستوى معنوية ٠.٠١.

الحل:

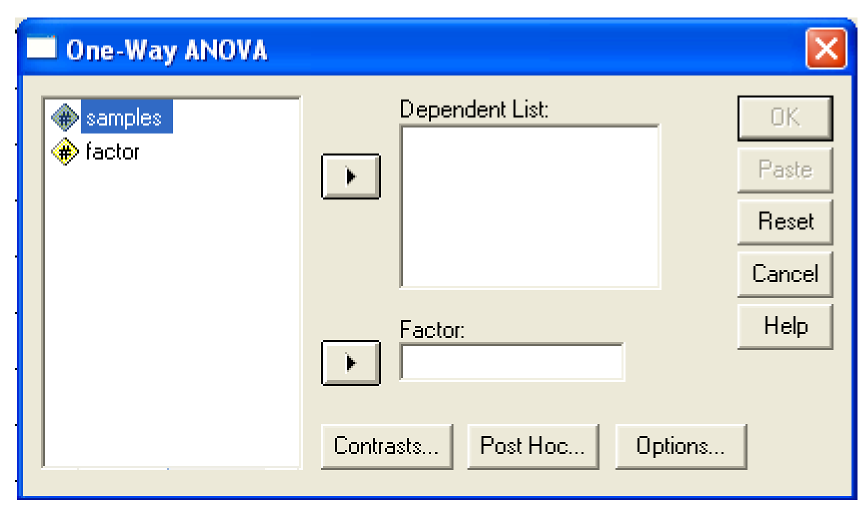
في البداية يتم إدخال البيانات كلها في البرنامج في عمود واحد مع إضافة عمود ثاني يشير إلى رقم العينة فيكون شاشة data editor وقد سمناه factor نضع فيه قيمة ١ لرمز أن القيمة الموجودة في العمود samples هي قيمة العينة قيمة الدواء الأولى a ونضع القيمة ٢ لرمز أن القيمة الموجودة في العمود samples هي قيمة الدواء الثاني b وهكذا ونضع القيمة ٣ لرمز أن القيمة الموجودة في العمود samples هي قيمة الدواء الثالث c وهكذا ونضع القيمة ٤ لرمز أن القيمة الموجودة في العمود samples هي قيمة الدواء الرابع d وتكون البيانات كما في الشكل التالي:

samples	factor
7.00	1
3.00	1
4.00	1
6.00	1
5.00	1
6.00	2
5.00	2
9.00	2
8.00	2
7.00	2
3.00	3
2.00	3
5.00	3
6.00	3
4.00	3
8.00	4
6.00	4
3.00	4
7.00	4
6.00	4

وبعد إدخال البيانات نتبع الخطوات التالية:

- اختيار Compare means.
- اختيار One-Way ANOVA.

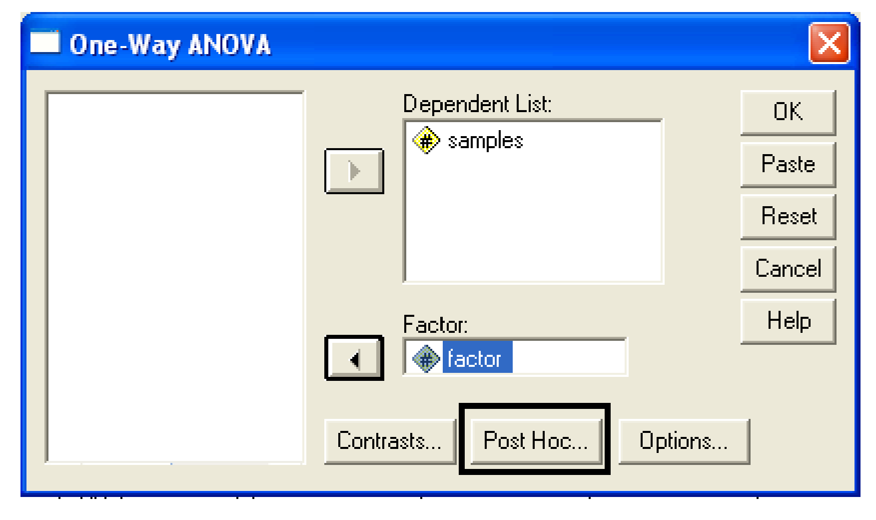
فيظهر الشكل الاتي:



شكل (٥)

⌖ ننقل المتغيرات المراد حساب الفروق فيها إلـى Dependent List، ومتغير تصنيف المشاهدات في Factor.

فيظهر الشكل الآتي:



شكل (٦)

⌖ ننقر على OK تظهر النتائج كما في الجدول التالي:

جدول (٧)

ANOVA

SAMPLES

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	25.000	3	8.333	3.030	.060
Within Groups	44.000	16	2.750		
Total	69.000	19			

القرار الإحصائي:

يتضح من الجدول أن p-value = 0.06 وهـو أكبر من 0.01 عند F = 3.03 وهذا يـدل علـى عدم وجود فروق بين الأنـواع الأربـع من الأدويـة ويؤكد ذلك جدول تحليل الفروق التالية:

Post Hoc Tests

جدول (8)

Multiple Comparisons

Dependent Variable: SAMPLES
LSD

(I) FACTOR	(J) FACTOR	Mean Difference (I-J)	Std. Error	Sig.	99% Confidence Interva Lower Bound	Upper Bound
1	2	-2.0000	1.049	.075	-5.0633	1.0633
	3	1.0000	1.049	.355	-2.0633	4.0633
	4	-1.0000	1.049	.355	-4.0633	2.0633
2	1	2.0000	1.049	.075	-1.0633	5.0633
	3	3.0000	1.049	.011	-6.334E-02	6.0633
	4	1.0000	1.049	.355	-2.0633	4.0633
3	1	-1.0000	1.049	.355	-4.0633	2.0633
	2	-3.0000	1.049	.011	-6.0633	6.334E-02
	4	-2.0000	1.049	.075	-5.0633	1.0633
4	1	1.0000	1.049	.355	-2.0633	4.0633
	2	-1.0000	1.049	.355	-4.0633	2.0633
	3	2.0000	1.049	.075	-1.0633	5.0633

نلاحظ أنه باستعمال طريقة LSD لم تظهر فروق معنوية بمستوى معنوية (دلالة) 0.01 بين متوسطات المعالجات كافة حيث كانت قيمة P-Value أو .Sig أكبر من 0.05 مع عدم وجود علامة (*) عند فروق المتوسطات .

تحليل التباين في اتجاهين (الثنائي) (Two Way ANOVA):

إن تحليل التباين في اتجاه واحد يستخدم لدراسة أثرعامل واحد فقط على متغير ما ولكن عندما نريد دراسة أثر عاملين أو أكثر في هذه الحالة يمكننا استخدام تحليل التباين في اتجاهين (أو ما يسمى تحليل التباين الثنائي).

فتحليل التباين الثنائي Two Way ANOVA يمكن استخدامه لدراسة أثر متغيرين عاملين نسمي أحدهما الصفوف (المعالجات)، والآخر الأعمدة (القطاعات) ويمكن وصف البيانات تحت هذه التجارب كالتالي:

جدول (9)

الصفوف (Rows)		الأعمدة (Columns) القطـــــاعــات					
		1	2		j	…	c
	1	X_{11}	X_{12}	...	X_{1j}	...	X_{1c}
	2	X_{21}	X_{22}	...	X_{2j}	...	X_{2c}
	.	.	.	...	.	...	.
	.	.	.	...	.	...	.
	.						
معالجات	i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ic}
	.	.	.	...	.	...	.
	.						
	.	.	.	...	.	...	.
	.	.	.	...	.	...	.
	.	.	.	…	.	…	.
	r	X_{r1}	X_{r2}	...	X_{rj}	....	X_{rc}

⌖ يمكن صياغة الفروض الإحصائية في هذه الحالة كالتالي:

(أ) **للمعالجات:**

$$H_0^{(1)} : \mu_1 = \mu_2 = \ldots = \mu_r$$

ضد الفرض البديل

$H_A^{(1)}$:يوجد متوسطان على الأقل غير متساوين

(ب) **للقطاعات:**

$$H_0^{(2)} : \mu_1 = \mu_2 = \ldots = \mu_c$$

ضد الفرض البديل

$H_A^{(2)}$:يوجد متوسطان على الأقل غير متساوين

مثال (٣)

القطاعات/ المعالجات	القطاع الأول	القطاع الثاني	القطاع الثالث	القطاع الرابع	المجموع
A	9.3	9.4	9.6	10	38.3
B	9.4	9.3	9.8	9.9	38.4
C	9.2	9.4	9.5	9.7	37.8
D	9.7	9.6	10	10.2	39.5
المجموع	37.6	37.7	38.9	39.8	154

إذا كان لدينا أربعة أنواع من الأسمدة D،C،B،A استخدمت كمعالجات لأربع قطاعات مختلفة، حيث كانت النتائج كما في الجدول التالي:

استخدم تحليل التباين في اتجاهين لمعرفة ما إذا كان لها تأثير مختلف على زيادة إنتاج القمح أم لا؟

الحل:

أولا الفروض الإحصائية بالنسبة للمعالجات:

H_0 : لا يوجد فروق بين متوسطات المعالجات (الصفوف).

H_1 : يوجد فروق بين متوسطات المعالجات على الأقل لاثنين منهم.

ثانيا الفروض الإحصائية بالنسبة للقطاعات:

H_0 : لا يوجد فروق بين متوسطات القطاعات (الأعمدة).

H_1 : يوجد فروق بين متوسطات القطاعات على الأقل لاثنين منهم.

ندخل البيانات كما في الشكل التالي:

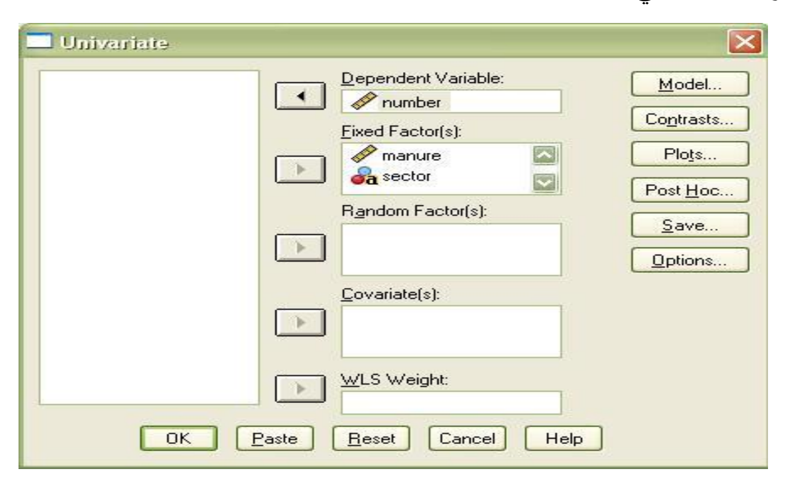
	manure	sector	number
1	1.00	A	9.30
2	1.00	B	9.40
3	1.00	C	9.20
4	1.00	D	9.70
5	2.00	A	9.40
6	2.00	B	9.30
7	2.00	C	9.40
8	2.00	D	9.60
9	3.00	A	9.60
10	3.00	B	9.80
11	3.00	C	9.50
12	3.00	D	10.00
13	4.00	A	10.00
14	4.00	B	9.90
15	4.00	C	9.70
16	4.00	D	10.20

الشكل (٧)

ثم نتبع الآتي:

Analyze → General linear Model → Univariate

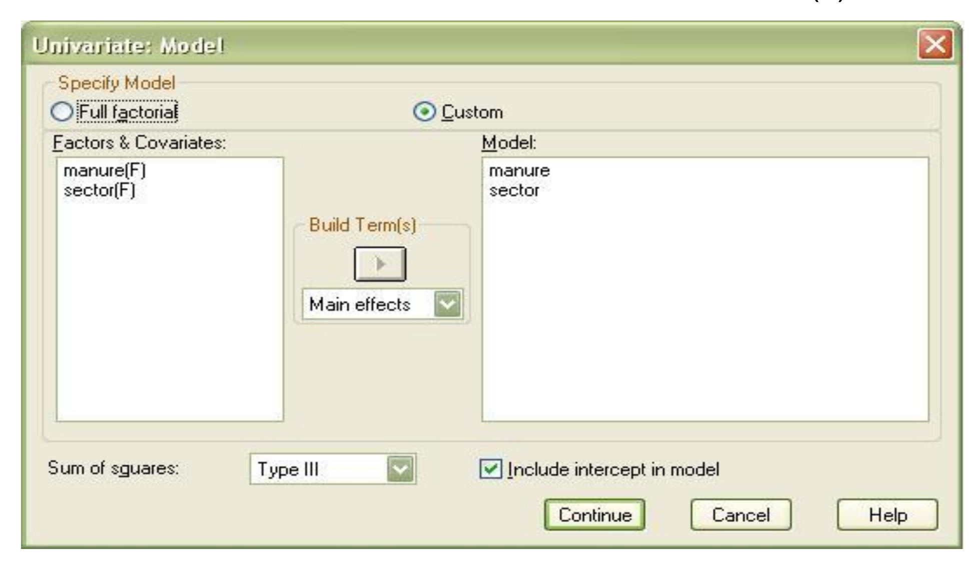
ليظهر الشكل التالي:



الشكل (٨)

- أدخلنا العاملين في Fixed Factors على اعتبار أن كافة معالجات العامل قد ضمنت في التجربة أما في حالة أخذ عينة من معالجات العامل فإننا سنستعمل خانة Random Factor.
- انقر زر Model فيظهر صندوق حوار Model، اختر Custom بدلا Full Factorial وذلك لأننا لا نرغب في ظهور التفاعل Interaction في جدول

تحليل التبـاين (Manure*Sector) لعدم وجـود درجـات حريـة كافيـة للخطـأ التجريبـي حيـث يظهـر صـندوق حـوار Model بعـد ترتيبـه كمـا فـي الشـكل (٩):



الشكل (٩)

حيث أن تأشير Include Intercept In Model يعمل علـى تضـمين الحـد الثابـت فـي النموذج الخطي العام باعتباره نموذج انحدار.

أما خانـة Build Terms فتستعمل لتعيين نـوع التـأثيرات Effects التـي يـراد إظهارها فــي جـدول تحليـل التبــاين حيـث قمنــا بنقـل المتغيـرين manure و sector مـن خانــة Factors & Covariates إلى خانة Model، لتظهر كتـأثيرات رئيسية Main Effect (بعد التأكد من أن خانـة Build Effects يتضمن الخيـار Main effect ثم انقر علـى Continue ثم Ok في صندوق حوار Univariate.

ثم تحصل على المخرجات التالية كما في الشكل التالي:

جدول (١٠)

Between-Subjects Factors

		N
manure	1.00	4
	2.00	4
	3.00	4
	4.00	4
sector	A	4
	B	4
	C	4
	D	4

Tests of Between-Subjects Effects

Dependent Variable: number

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1.210[a]	6	.202	22.687	.000
Intercept	1482.250	1	1482.250	166753.1	.000
manure	.825	3	.275	30.937	.000
sector	.385	3	.128	14.437	.001
Error	.080	9	.009		
Total	1483.540	16			
Corrected Total	1.290	15			

a. R Squared = .938 (Adjusted R Squared = .897)

الجدول الأول يعطي بيان بعدد الحالات و المفقود، الجدول الثاني يبين مجموع المربعات بين المعالجات = 0.385 و مجموع المربعات بين القطاعات = 0.825 مجموع مربعات الخطأ = 0.08 و مجموع المربعات الكلي = 1.29.

- **القرار الإحصائي للمعالجات:**

و أيضا يتضح معنوية الفرق بين الأسمدة وكذلك بين القطـاعات بـمستوى معنويـة 0.05 حيث أن

p-value=0.0 عند صف الأسمدة manure وهي أقل من 0.05 أي نـرفض H_0 ونقبل H_A ونقول أنه يوجد فروق بين متوسطات المعالجات .

- **القرار الإحصائي للقطاعات:**

وأيضا p-value=0.001 عند صف القطاعات sector وهـي أقل من 0.05 نـرفض H_0 ونقبل H_A ونقول أنه يوجد فروق بين متوسطات القطاعات.

الفصل السادس

تحليل الارتباط

Correlation Analysis

تعرضنا في الفصول السابقة لدراسة ظاهرة واحدة كالاستهلاك، الدخل، المبيعات، العمر، الوزن وبدأنا ببيان كيفية عرض البيانات التي جمعناها في جدول أو أشكال بيانية مختلفة واستطردنا بعد ذلك إلى كيفية تلخيص تلك البيانات من خلال مقاييس النزعة والتشتت لمجموعة واحدة أو مجموعات مختلفة وكيفية تقدير تلك المقادير وكيفية اختبارات الفروض عليها و الآن نبحث كيفية إيجاد العلاقات الرياضية التي تربط المتغيرات ببعضها وما مقدار هذا الارتباط من خلال ما يسمي بمقاييس الارتباط المختلفة والتي سوف نتعرض لها في هذا الفصل كما نناقش في الفصل التالي كيف يمكن التنبؤ بأحد المتغيرات لقيمة محددة للمتغير الأخر و هي ما تسمي معادلة الانحدار.

نستطيع القول بأن تحليل الارتباط والانحدار هو أداة إحصائية نستفيد منها في تحديد العلاقة بين متغيرين أو أكثر للتنبؤ بأحد المتغيرات استنادا إلى قيم المتغير أو المتغيرات الأخرى. فمثلا إذا علمنا العلاقة بين مصروفات الدعاية و المبيعات، فيمكننا الاستفادة من تحليل الارتباط للتنبؤ بالمبيعات حالما تتوفر لنا قيمة نفقات الدعاية.

وسوف نتناول في هذا الفصل قضية الارتباط بين متغيرين سواء كانوا كميين أو غير ذلك بينما في الفصل القادم نتناول كيفية التنبؤ بأحد المتغيرات في ضوء متغير آخر أو أكثر من خلال دارسة الانحدار

الارتباط الخطى البسيط Simple Correlation

تسمي العلاقة بين ظاهرتين بالارتباط Correlation مثلاً العلاقة بين الدخل والاستهلاك فمن البديهي أن زيادة دخل الفرد يؤدي إلى زيادة استهلاكه من السلع والخدمات (علاقة طردية) كما أن ارتفاع سعر سلعة ما يؤدي إلى تدني الطلب عليها (علاقة عكسية) علماً أن الارتباط قد يكون خطياً Linear أو غير خطي Non Linear.
إن المقياس المستخدم الذي يقيس درجة الارتباط يعرف بمعامل الارتباط Correlation Coefficients ويرمز له r وتتراوح قيمته بين -١ إلى ١ $[-1 \le r \le 1]$.

ومن الأمثلة على ذلك:

- الإنفاق، والدخل العائلي.
- سعر السلعة، والكمية المطلوبة منها.
- الفترة الزمنية لتخزين الخبز، وعمق طراوة الخبز.
- تقديرات الطلاب في مقرر الإحصاء، وتقديراتهم في مقرر الرياضيات.

- كميات السماد المستخدمة، وكمية الإنتاج من محصول معين تم تسميده بهذا النوع من السماد.
- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.
- وزن الجسم، وضغط الدم.

يحسب معامل الارتباط الخطي البسيط بافتراض وجود علاقة خطية بين اثنين من المتغيرات فقط مع العلم أن الحصول على قيمة صغيرة (قريبة من الصفر) لهذا المعامل لا يعني عدم وجود علاقة بين المتغيرين فقد توجد علاقة من الدرجة الثانية (ارتباط غير خطي). ويختلف نوع المقاييس الذي نستخدمه في حساب معامل الارتباط طبقا لنوع البيانات وسوف يجرى حسابه في حالة البيانات الكمية ، والبيانات الوصفية المقاسة بمعيار ترتيبي.

- **الغرض من تحليل الارتباط الخطى البسيط**

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (وتقرأ "رُو")، وفي حالة العينة بالرمز r ، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

- نوع العلاقة:ـ وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

✓ إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس.

✓ إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس .

✓ إذا كان قيمة معامل الارتباط صفرا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

قوة العلاقة:-

يمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن $[\pm 1]$، حيث أن قيمة معامل الارتباط تقع في المدى $[-1 \le r \le 1]$.

⌖ معامل الارتباط الخطى البسيط " لبيرسون" Karl Pearson

في حالة جمع بيانات عن متغيرين كميين $(y\ ,\ x)$، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" *Pearson*، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة بين الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

إذا افترضنا أن لدينا عينة مكونة من n من أزواج المشاهدات $(x_1\ ,y_1),\ (x_2,y_2),\ ....,(x_n\ ,y_n)$ لمتغيرين X و Y، فإن معامل الارتباط لبيرسون يعرف على أنه:

$$r = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} Y_i^2 - \frac{\left(\sum_{i=1}^{n} Y_i\right)^2}{n}\right]}} \qquad (١)$$

فإذا نظرنا إلى المعادلة السابقة عرفنا مقدار المعاناة عند حساب مقدار معامل الارتباط ناهيك عن حجم التحليل والحسابات إذا زادت حجم العينة و الآن نتعرف كيف يمكن تطبيق ذلك بمنتهى السهولة باستخدام البرنامج وذلك من خلال المثال القادم:

مثال(١):

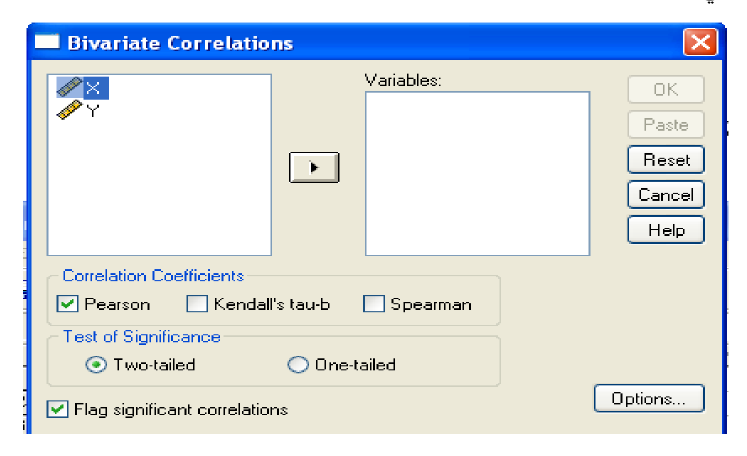
احسب معامل الارتباط المناسب بين درجة الطالب في الرياضيات ودرجته في الإحصاء باستخدام البيانات الآتية لعينة من خمسة طلاب:

جدول (١)

درجة الرياضة (x)	١٥	١٩	١١	٣	٧
درجة الإحصاء (y)	١٨	٢٠	١٦	١٤	١٢

الحل:

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Bivariate فيظهر الشكل الآتي:



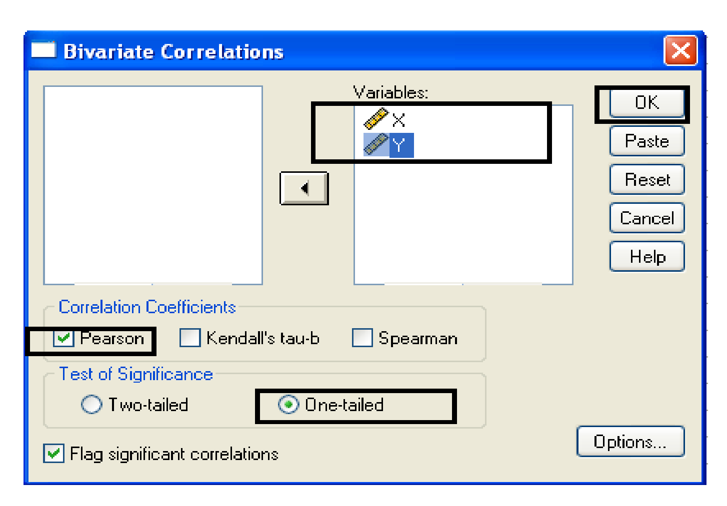
شكل (١)

ونرى أن الشكل السابق يتضمن مجموعة من الأجزاء:

- عمود المتغيرات المتاحة وهو العمود الأول من جهة اليسار ومنه يتم نقل المتغيرات إلى العمود الذي بجواره لنبدأ حساب معامل الارتباط وهنا نقوم بنقل X،Y .
- Correlation Coefficients: وهو يتيح لنا تحديد نوع معامل الارتباط الذي نريده ونلاحظ أنه متوفر لدينا ٣ اختيارات:
  - ✓ Pearson: وهو معامل ارتباط الظواهر الكمية سواء كانت مبوبة أو غير مبوبة.
  - ✓ Kendall's tau: وهو معامل ارتباط الظواهر الترتيبية ويفضل للعينات الصغيرة والبيانات الغير مبوبة .
  - ✓ Spearman: وهو أيضا معامل الارتباط بين الظواهر الكمية والترتيبية ويفضل للعينات الصغيرة ويستخدم في حالة البيانات الغير مبوبة.
  - ✓ وهنا نقوم بتحديد معامل بيرسون.
  - ✓ Test of Siginificance: ولأننا نستخدم بيانات عينة فإننا نحتاج إلى التأكد من قيمة معامل الارتباط فنحتاج إلى اختبار هذة القيمة وهنا نختار إما الاختبار من جهة أو جهتين.

✓ Flag Significant Correlation: هنا تستخدم لتعليم الارتباط بعلامة نجمة (star) في حالة الارتباط المعنوي (الدال) أي الارتباط الحقيقي والذي يؤكده الاختبار.

وبعد الاختيار السابق يظهر الشكل (٣) في الشكل الآتي:



شكل (٢)

⌖ ثم ننقر على ok فيظهر الجدول الآتي:

جدول (٢)

Correlations

		X	Y
X	Pearson Correlation	1	.900*
	Sig. (1-tailed)		.019
	N	5	5
Y	Pearson Correlation	.900*	1
	Sig. (1-tailed)	.019	
	N	5	5

*. Correlation is significant at the 0.05 level (1-tailed).

⌖ ونجد أن معامل الارتباط بين درجة الطالب في الإحصاء ودرجته في الرياضيات مقدارها 0.9 وهذا يعني أن الارتباط طردي وقوي جدا ولاختبار الفرض الصفري:

$$H_0 : \rho = 0$$

ضد الفرض البديل:

$$H_A : \rho \neq 0.$$

ونلاحظ أن الفرض البديل لا يساوي صفر وهذا ما دل عليه اختيارنا السابق أن الاختبار ذو اتجاهين فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقدارها $\alpha = 0.05$ فإن قيمة p-value = $\alpha = 0.05$ >0.019 دل ذلك على أن الارتباط بين درجة الرياضيات و الإحصاء ارتباط يختلف عن الصفر بمستوى معنوي مقداره ٠.٠٥ أي أننا نقبل الفرض البديل القائل بأن الارتباط لا يساوي صفر.

- **معامل ارتباط الرتب (لسبيرمان) Spearman**

إذا كانت الظاهرة محل الدراسة تحتوي على متغيرين وصفيين ترتيبين، ومثال على ذلك قياس العلاقة بين تقديرات الطلبة في مادتين ، أو العلاقة بين درجة تفضيل المستهلك لسلعة معينة ، ومستوى الدخل، فإنه يمكن استخدام طريقة "بيرسون" السابقة في حساب معامل ارتباط يعتمد على رتب مستويات المتغيرين كبديل للقيم الأصلية ، ويطلق على هذا المعامل " معامل ارتباط سبيرمان " Spearman.

مثـــال (٢):

فيما يلي بيانات درجة تفضيل ومستوى دخل لعشرة مستهلكين لسلعة ما كعينة والمطلوب معرفة الارتباط بين مستوى الدخل ودرجة تفضيل المستهلكين للسلعة وما مدلول ذلك بالنسبة لجميع المستهلكين:

جدول (٣)

مستوي التفضيل	أفضل	لا أفضل	أفضل	لا أفضل	لا أفضل	لا أفضل	أفضل	لا أفضل	أفضل	أفضل
مقدار الدخل الشهري	١٠٠	٢٠٠	٢٥٠	٤٠٠	١٣٠	٢٧٠	٧٠٠	٥٠٠	٣٣٠	٢٠٠

الحـــل:

خلال البرنامج نقوم بإدخال البيانات على الفرض أن "أفضل" نرمز لها بالرمز ١ و"لا أفضل" برمز ٢.

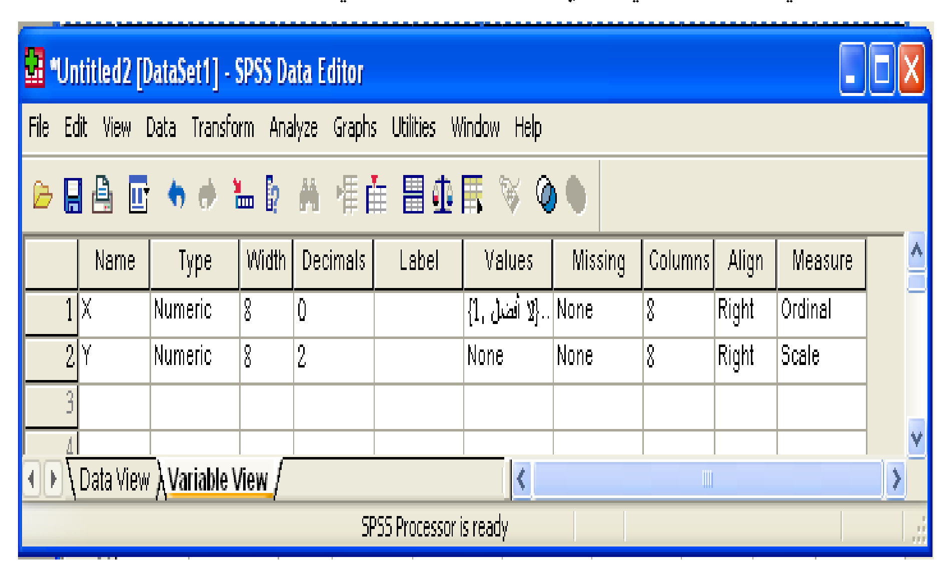
ونقوم بتعريف المتغير الترتيبي x كما يلي في محرر المتغيرات نقوم بتعريف قيم المتغير في عمود values كما يلي:

Value Labels

Value Labels
Value:
Label:
Add
Change
Remove
1 = "افضل"
2 = "لا افضل"
OK
Cancel
Help

شكل (٣)

ثم في عمود measure نعرف المتغير X على أنه متغير ترتيبي ordinal ونعرف المتغير Y على أنه متغير كمي عادي فيظهر المحرر كالآتي:



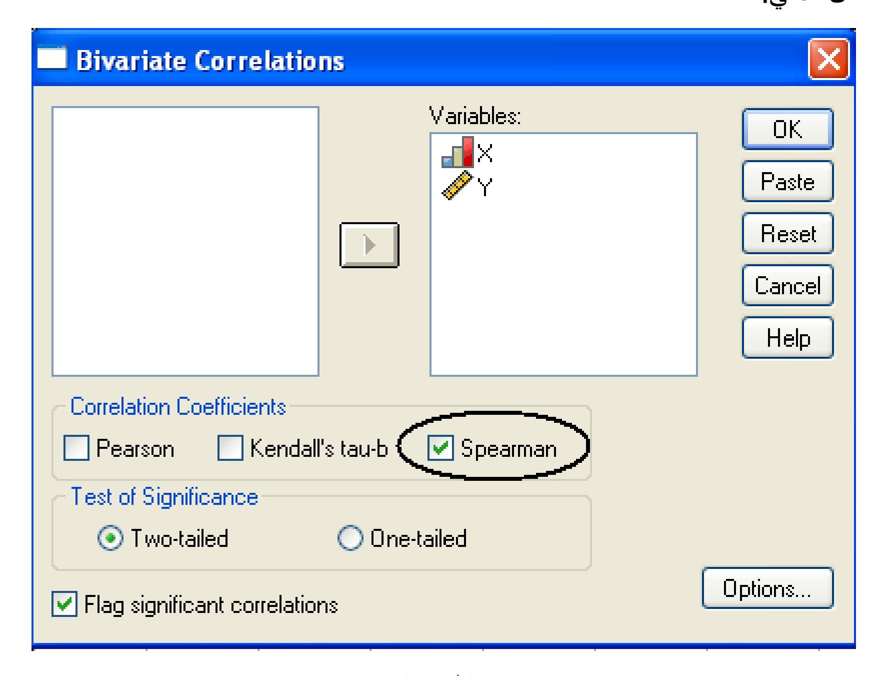
شكل (٤)

ثم شاشة محرر البيانات نقوم بإدخال البيانات السابقة فتظهر في الشكل الآتي:

	X	Y
1	1	100.00
2	2	200.00
3	1	250.00
4	2	400.00
5	2	130.00
6	2	270.00
7	1	700.00
8	2	500.00
9	1	330.00
10	1	200.00

شكل (٥)

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Bivariate فيظهر الشكل الآتي:



شكل (٦)

ونلاحظ أن المتغير x معرف في عمود المتغير المستخدمة في التحليل (العمود الأول من اليمين) وأننا اخترنا spearman كمعامل للارتباط ونفس الاختيارات السابقة في مقياس Pearson ثم ننقر على ok فيظهر الجدول الآتي:

جدول (٤)

Correlations

			X	Y
Spearman's rho	X	Correlation Coefficier	1.000	.070
		Sig. (2-tailed)	.	.848
		N	10	10
	Y	Correlation Coefficier	.070	1.000
		Sig. (2-tailed)	.848	.
		N	10	10

ونجد أن معامل الارتباط بين درجة تفضيل المستهلكين لسلعة ومستوى دخولهم مقداره 0.070 وهذا يعني أن الارتباط طردي وضعيف ولاختبار الفرض القائل:

$$H_0 : \rho = 0$$

ضد الفرض البديل القائل :

$$H_A : \rho \neq 0.$$

ونلاحظ أن الفرض البديل معامل الارتباط لا يساوي صفر وهذا ما دل عليه اختيارنا السابق أن الاختبار ذو اتجاهين فإذا افترضنا أننا نريد اختبار ذلك بمستوى معنوية مقدارها $\alpha = 0.05$ فإن قيمة p-value = $\alpha = 0.05$ <0.424 دل ذلك على أن الارتباط بين تفضيل المستهلكين لسلعة ومستوى دخولهم ارتباط يساوي الصفر بمستوى معنوي مقداره ٠.٠٥ أي أننا نقبل الفرض العدم القائل الارتباط يساوي صفر في كامل بيانات المجتمع.

الارتباط الجزئي Partial Correlation

يقيس معامل الارتباط الجزئي قوة العلاقة بين متغيرين بثبوت متغير ثالث أو أكثر. مثلا قد نحصل على قيمة عالية لمعامل الارتباط البسيط للعلاقة بين أسعار اللحوم البيضاء واللحوم الحمراء فقد لا توجد علاقة فعلية بين المتغيرين ولكن كلا المتغيرين يتأثر بمعامل ثالث هو المستوى العام للأسعار فإذا استبعدنا المستوى العام للأسعار (أو تثبيته) عند قياس العلاقة بين أسعار اللحوم البيضاء والحمراء فسيتم الحصول على قيمة أقل لمعامل الارتباط وهذا يعرف بالارتباط الجزئي . علما أنه يمكن استبعاد أي عدد من المتغيرات عند قياس العلاقة بين ظاهرتين.

والآتي مثال يوضح كيفية تطبيق ذلك من خلال البرنامج:

مثـــال (٣):

البيانات الآتية تعطي الدخل الشهري لمجموعة أسر (X) والإنفاق الشهري على الطعام (Y) وحجم مدخراتها الشهرية (Z) وذلك في عينة من تسعة أسر:

جدول (٥)

الأسرة	١	٢	٣	٤	٥	٦	٧	٨	٩
x	٤٠	٦٥	٩٠	٦٠	٧٥	٨٠	١٢٠	٤٥	٦٥
Y	٢٢	٢٨	٤٠	٢٠	٣٢	٣٢	٥٠	٢٠	٣٥
Z	٦	١٠	٨	٢٥	١٢	٧	١٠	٥	٧

المطلوب:

١- حساب معامل الارتباط الجزئي لـ x و y بثوت z.

٢- حساب معامل الارتباط الجزئي لـ x و z بثوت y.

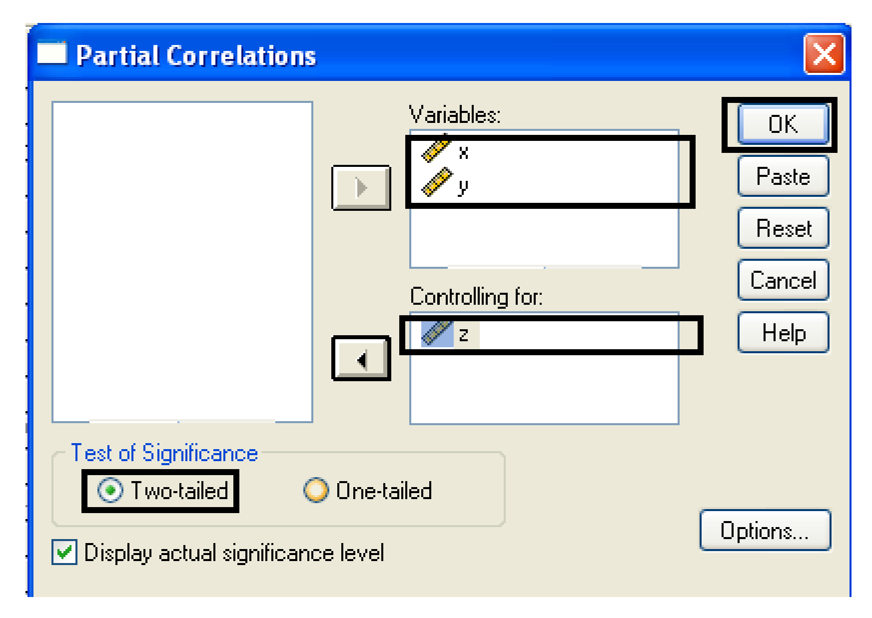
٣- حساب معامل الارتباط الجزئي لـ z و y بثوت x.

٤- اختبار معنوية (دلالة) الارتباط بمستوى معنوية قدره ٠.٠٥.

الحل:

أولا: حساب معامل الارتباط الجزئي لـ x و y بثوت z.

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Partial فيظهر مربع حوار معامل الارتباط الجزئي كما في الشكل التالي:



شكل (٧)

في خانة variables يتم إدخال المتغيرات التي يراد حساب معامل الارتباط الجزئي لها x، y ، وفي خانة Controlling for يتم إدخال المتغير (المتغيرات) الذي يراد استبعاد أثره (z).

وعند النقر على زر ok نحصل على النتيجة التالية:

جدول (٦)

Correlations

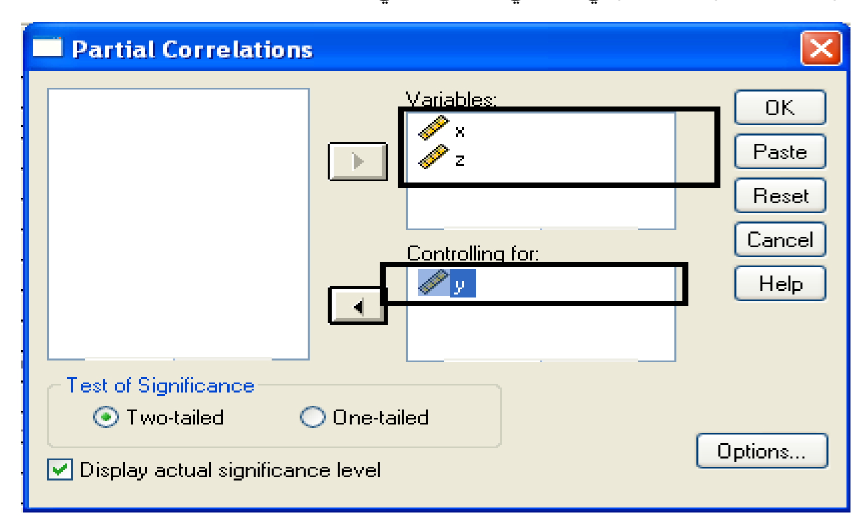
Control Variables			x	y
z	x	Correlation	1.000	.965
		Signif icance (2-tailed)	.	.000
		df	0	6
	y	Correlation	.965	1.000
		Signif icance (2-tailed)	.000	.
		df	6	0

نلاحظ أن قيمة معامل الارتباط من الشكل السابق تساوي ٠.٩٦٥ أي أنه توجد علاقة طردية وقوية بين الدخل والإنفاق وكما أن قيمة p-value=0 وهي أقل من قيمة

مستوى المعنوية مقسوم على ٢=٠.٠٢٥ حيث أننا نختبر فرض عدم ذو اتجاهين ولذلك فإننا نقبل الفرض البديل القائل أن الارتباط بين المتغيرات في المجتمع لا يساوي صفر ٠

ثانياً: حساب معامل الارتباط الجزئي لـ x و z بثوت y.

من شريط قوائم نختار Analyze ثم نختار Correlate ثم نختار Partial فيظهر مربع حوار معامل الارتباط الجزئي كما في الشكل التالي:



شكل (٨)

في خانة variables يتم إدخال المتغيرات (x، z) ، وفي خانة Controlling for يتم إدخال المتغير (y). عند النقر على زر ok نحصل على النتيجة التالية:

جدول (٧)

Correlations

Control Variables			x	z
y	x	Correlation	1.000	.693
		Signif icance (2-tailed)	.	.057
		df	0	6
	z	Correlation	.693	1.000
		Signif icance (2-tailed)	.057	.
		df	6	0

إن قيمة معامل الارتباط من الشكل السابق تساوي ٠.٦٩٣ أي أنه توجد علاقة طردية بين الدخل الادخار وكما أن قيمة p-value=٠.٠٥٧ وهي أكبر من قيمة مستوى المعنوية

مقسوم على ٢ حيث أننا نختبر فرض عدم ذو اتجاهين ولذلك فإننا نقبل الفرض العدم القائل أن الارتباط بين المتغيرات في المجتمع يساوي صفر.

ثالثاً: حساب معامل الارتباط الجزئي لـ z و y بثوت x.

كما فعلنا في أولا وثانيا في إدخال وتعريف المتغيرات نفعل في تلك الخطوة وتظهر نتيجة معامل الارتباط كآلاتي:

جدول (٨)

Correlations

Control Variables			z	y
x	z	Correlation	1.000	-.710
		Signif icance (2-tailed)	.	.048
		df	0	6
	y	Correlation	-.710	1.000
		Signif icance (2-tailed)	.048	.
		df	6	0

فنجد أن قيمة معامل الارتباط بين الاستهلاك والإدخار تساوي ٠.٧١٠ - وهذا يعني أنه توجد علاقة بين الاستهلاك والإدخار عكسية وقوية، ونجد أيضا أن قيمة p-value=٠.٠٤٨ وهي أكبر من قيمة مستوى المعنوية مقسوم على ٢=٠.٠٢٥ حيث أننا نختبر فرض عدم ذو اتجاهين ولذلك فإننا نقبل الفرض العدم القائل أن الارتباط بين المتغيرات في المجتمع يساوي صفر في ضوء بيانات العينة.

الفصل السابع

الاختبارات اللامعلمية (اللابارامترية)

Non-parametric Tests

في معظم الأساليب التي تكلمنا عنها في الاختبارات المعلمية (البلارامترية) نجد أنها مبنية على الفرضية التي تقول بأن العينة أو العينات العشوائية التي تم اختيارها للدراسة من مجتمع طبيعي، وغالبا ما تكون هذه الأساليب غير دقيقة إلى حد ما عندما يكون مجتمع العينة غير طبيعي، وحيث أن بعض المجتمعات لا تفي بالشروط المطلوبه لتطبيق تلك الأساليب، دعت الحاجة للبحث عن أساليب أخرى لا يتطلب تطبيقها مثل ذلك الشرط. هذه الأساليب يطلق عليها تسمية الأساليب اللامعلمية (اللابارامترية) لأنه وكما لوحظ في الفصل الأول كان اهتمامنا يركز على معلمة (بارامتر) أو أكثر من معلمات (بارامترات) المجتمع الإحصائي (المتوسط، التباين، النسبة،...الخ) علاوة على ذلك، وكما اشرنا في الفصل الأول، لكي نصل إلى استنتاج إحصائي يجب معرفة صيغة التوزيع الاحتمالي للمجتمع التي تم اختيار العينة منها.

وهناك نوعان من الأساليب الإحصائية تتم معاملتها على أنها أساليب لا معلمية وهما: أساليب لا معلمية بما تعنيه الكلمة وهي أساليب تختبر الفرضيات التي لا تتضمن أي نص يتعلق بمعلمات المجتمع الإحصائي، أما الأساليب الأخرى فهي أساليب التوزيعات الحرة وهي الأساليب التي لا تضع أي افتراضات على مجتمع العينة، و بصرف النظر عن التميز بين هذين الأسلوبين فإن كلاهما ستتم معاملتها على أنهما أساليب لا معلمية، هذه الأساليب يتم تطبيقها على سبيل المثال لا الحصر في الحالات التالية:

١-إذا كانت الفرضية المطلوب اختبارها لا تتضمن معلمة المجتمع.

٢-البيانات مقاسه بمقياس أضعف من المقاييس المطلوبة لتطبيق الأساليب المعلمية مثل (المقياس الاسمي، المقياس الترتيبي، مقياس الفترة، المقياس النسبي).

٣-إن لم تتوفر الشروط المطلوبة لتطبيق الأساليب المعلمية.

بعض مزايا الاختبارات اللامعلمية (اللابارامترية) ما يلي:

- مهما كان شكل التوزيع المأخوذ منه العينة فإن الاختبار اللامعلمي الذي له مستوى معنوية (دلالة) معين يكون له هذا المستوى فعلاً بشرط أن تكون العينة قد اختيرت عشوائياً كما يشترط أيضاً في بعض الحالات استمرار التوزيع.
- الإحصاءات اللامعلمية هي الأسلوب الوحيد الممكن استخدامه في حالة العينات الصغيرة جداً إلا إذا كان توزيع المجتمع معروفاً تماماً.

- يمكن استخدامها أحيانا للعينات التي تحتوي على مشاهدات من عدة مجتمعات متفاوتة.
- تصلح لتحليل البيانات التي تكون علي صورة رتب دون الحاجة إلى معرفة التوزيع في المجتمع الأصلي للبيانات.
- تستخدم في حالة كون البيانات تتضمن احدى صيغتي التفضيل مثلاً سليم أو معيب حيث السليم تكون له إشارة موجبة والمعيب إشارة سالبة وهنا لا تصلح الطرق التقليدية.
- يمكن تطبيقها عندما تكون البيانات مقاسه بمقياس ضعيف.
- تعتمد على افتراضات قليلة، وبالتالي فرصة تطبيقها خطأ ستكون صغيرة.
- الحسابات الضرورية للأساليب اللامعلمية عادة ما تكون سهلة ويمكن إنجازها بسرعة.
- سهولة فهمها وطريقة حسابها تجعلها مناسبة جدا للباحثين الذي ليس لهم خلفية علمية جيدة في الرياضيات و الإحصاء.

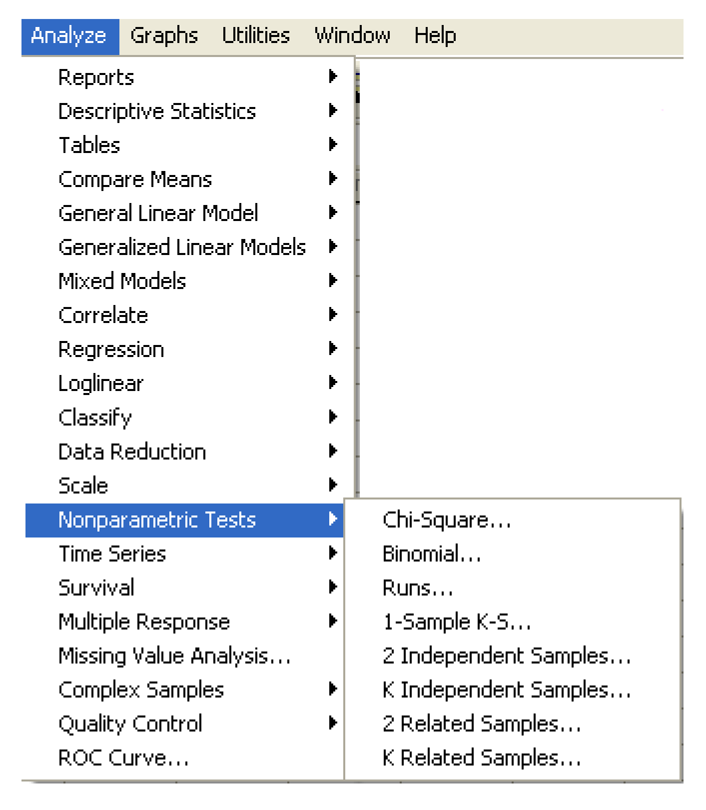
بعض عيوب الاختبارات اللامعلمية (اللابارامترية):

- نتيجة لسهولة حسابها، في بعض الأحيان يتم تطبيقها في مسائل يكون من الأفضل تطبيق أساليب معلمية عليها مما يسبب في ضياع المعلومات.
- في حالة العينات الكبيرة يؤدي استخدامها إلى جهد أكبر من الأساليب التقليدية.
- في حالة تحليل بيانات من توزيع طبيعي فإن استخدام الاختبارات اللامعلمية يعتبر فقد للبيانات وتقاس درجة الفقد بكفاءة الاختبار اللامعلمية.

وتنقسم الاختبارات اللامعلمية حسب عدد العينات عند إجراء الاختبار:

- حالة عينة واحدة One sample case
- حالة عينتين Two samples case (وهنا يوجد اختلاف بين المقياس التي تعتمد على: أ- العينتان مستقلتان ب- العينتان غير مستقلتين)
- حالة عدد العينات K والتي ربما تفترض استقلالا للعينات أو ارتباطها.

و تظهر أوامر الاختبارات اللامعلمية بالنقر على Analyze في شريط الأوامر فيظهر الشكل الآتي :



الشكل (١)

ومن الشكل السابق نستطيع اختيار نوع الاختبار الملائم لنوع المشكلة لدينا وسوف نتعرف فيما يلي على تلك الأنواع من الاختبارات وكيفية معالجتها للمشكلات المختلفة:

اختبار مربع كاي (Chi-Square test):

إن من أشهر وأقدم اختبارات جودة المطابقة هو اختبار مربع كاي لجودة المطابقة، الذي اقترحه بيرسون (١٩٠٠م)، ويستخدم هذا الاختبار لتحديد ما إذا كانت التكرارات المشاهدة في جدول توزيع تكراري بسيط (لظاهرة واحدة) تتبع توزيع احتمالي معين مثل:

- اختبار أن عدد الحوادث التي تقع في ميدان معين لها توزيع بواسون (وهو أحد التوزيعات الاحتمالية المتقطعة شائعة الاستخدام في كثير من التطبيقات)،

- اختبار أن درجات الطلبة في أحد الامتحانات لها توزيع طبيعي، اختبار أن متوسط أطوال القطع التي تنتجها احدى الآلات لها توزيع طبيعي.

في كل هذة الحالات وأمثالها يقوم الاختبار أساسا على مقارنة التكرارات المشاهدة بالتكرارات المتوقعة التي تحسب باستخدام ذلك التوزيع الاحتمالي المعين (المطلوب اختبار ما إذا كانت البيانات تتبعه أم لا) ومن الفروق بين التكرارات المشاهدة والمتوقعة تحسب χ^2 بتطبيق الصيغة التالية:

$$\chi^2 = \sum \frac{(observed - \exp ected)^2}{\exp ected}$$

Observed: القيم المشاهدة

Expected: القيم المتوقعة

و إذا كان الجدول التكراري به m من الخلايا تكون χ^2 المحسوبة لها توزيع χ^2 بدرجات حرية يساوي عدد الخلايا مطروحا منها ١ أي m-1 ومستوى معنوية مقداره α وعليه إذا كان

$$\chi^2_{calcuate} > \chi^2_{(\alpha, m-1)}$$

يرفض الفرض العدم عند مستوى معنوية α حيث فرض العدم هنا هو فرض أن التكرارات المشاهدة بالجدول تتبع التوزيع الاحتمالي المعين.

هذا الاختبار يشبه اختبارات كاي للاستقلالية والتجانس من حيث كون إن إحصاءة الاختبار تنتج من مقارنة التكرارات المشاهدة ولكن أوجه تطبيقها مختلف تماما.

- **شروط تطبيق الاختبار:**

✓ تتضمن البيانات عينة عشوائية بها n من المفردات المستقلة عن بعضها البعض تم اختيارها من مجتمع x، ويمكن وضع هذه البيانات في جدول توافقي كما يلي:

جدول (١)

الصنف	$1 \ 2 \ 3 ...... i ...... r$	المجموع
التكرار المشاهد	$O_1 \ O_2 \ O_3 ..... O_i ...... O_r$	n

حيث O_i تمثل عدد المفردات التي تقع في الصنف i حيث $i = 1,,2,3,...,r$ مع ملاحظة أنه من الممكن أن يكون التصنيف نوعي أو كمي، فمثلا من الممكن تصنيف مجموعة من الأشخاص حسب الجنس (ذكور، إناث) أو ممكن التصنيف بالعمر...الخ.

- ✓ وحدة القياس على الأقل اسمية (nominal).

الفروض الإحصائية:

إذا رمزنا لدالة التوزيع غير المعروفة لمجتمع X بالرمز F(x) و لدالة التوزيع الفرضية بالرمز $F_o(x)$ وهي محددة بالكامل عدا من الممكن أن تكون المعلمة غير معروفة، ويجب تقديرها من بيانات العينة، فإنه يمكن صياغة الفرضيات الإحصائية كما يلي:

$H_0 : F(x) = F_0(x)$ لجميع قيم x

$H_A : F(x) \neq F_0(x)$ على الأقل لقيمة واحدة من قيم x.

إحصاءة الاختبار:

حيث أنه هناك احتمال بأن تقع أي مفردة يتم اختيارها من المجتمع بأي صنف من التصنيفات المختلفة و بالتالي يمكن الرمز لهذه الاحتمالات بالرمز $p_1, p_2, ..., p_r$.

على التوالي وذلك لأنه يوجد r صنفا، وعليه في حالة H_0 يمكن حساب التكرار المتوقع بكل صنف كما يلي:

$$T = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i} \quad \& \quad E_i = np_i$$

حيث

O_i : القيم المشاهدة،

E_i: القيم المتوقعة،

P_i : قيمة الاحتمال،

n : مجموع المشاهدات.

القرار الإحصائي:

نرفض H_o إذا كان $T > \chi^2_{\alpha,r-1}$

مثال (١)

أخذت عينة عشوائية مكونة من ٤٠٠ أسرة من الأسر التي لكل منها ثلاثة أطفال أو أقل فوجد أن التوزيع التكراري لتلك الأسر حسب عدد الأطفال الذكور كالآتي:

جدول (٢)

عدد الأطفال الذكور x	٠	١	٢	٣
عدد الأسر	٤٨	١٤٧	١٥٣	٥٢

اختبر فرض أن عدد الأطفال الذكور بكل أسرة لديها ثلاثة أطفال له توزيع ذي الحدين بنجاح $\theta = 0.5$ وعدد المحاولات $N = 3$ وذلك عند مستوى معنوية مقدارها ٠.٠٥.

الحل:

- **أولا طريقة إدخال البيانات في البرنامج:**

نقوم بتصميم متغيرين أحدهم للمتغير الذي يصف عدد الأطفال في الأسرة و الآخر للمتغير الذي يصف التكرار المشاهد.

Number	observe
.00	48.0000
1.00	147.000
2.00	153.000
3.00	152.000

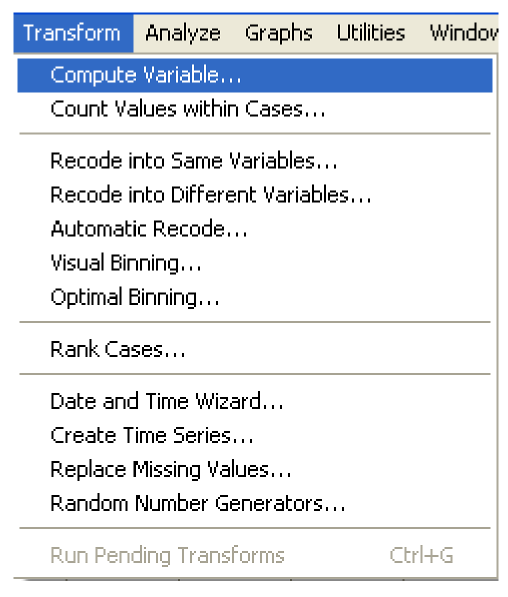
شكل (٢)

- **ثانياً: نقوم بحساب التكرار المتوقع :**

ونلاحظ أننا نريد حساب التكرار المتوقع من توزيع ذي لحدين ولذلك يتم أولا حساب الاحتمالات والذي نرمز لها بالرمز p_i ومن ثم حساب التكرار المتوقع كآلاتي:

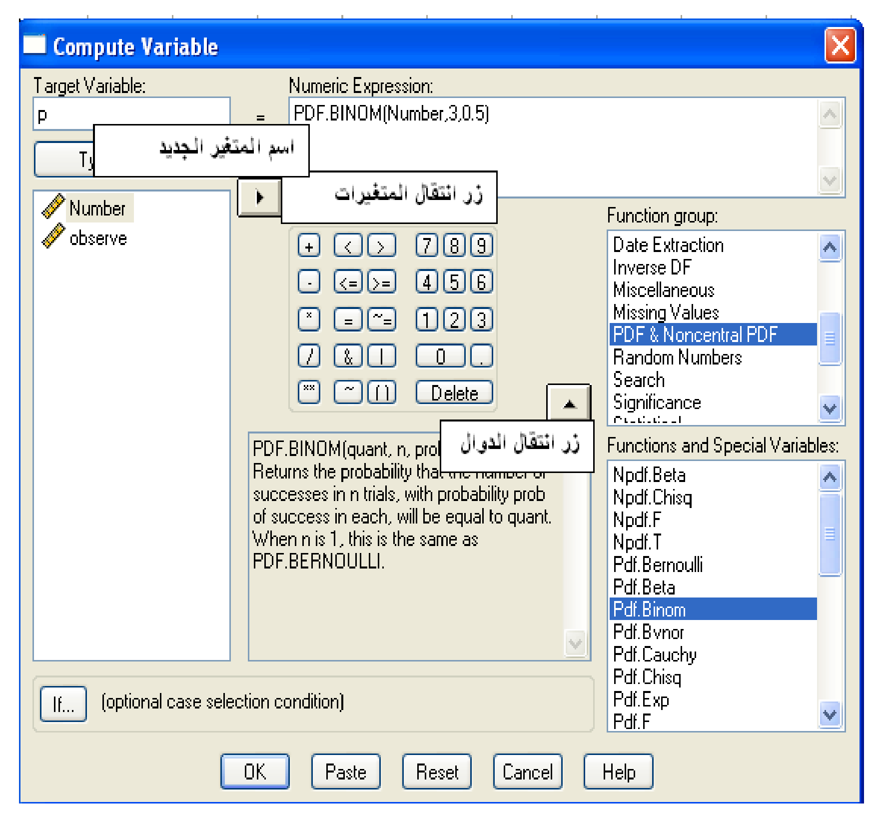
$$E_i = np_i$$
$$= 400\, p_i$$

فمن قائمة Transform نقوم باختيار ....Compute Variable ومن ثم يظهر الشكل الآتي:



شكل (٣)

ثم نقوم باختيار PDF and Noncentral PDF من عمود Function group وبعدها تظهر قائمة بدوال فرعية في عمود أسفل العمود السابق ونقوم باختيار pdf Binom وهذا يعني دالة الاحتمال لذي الحدين أي p_i فتضغط على الدالة مرتين متتالين أو النقر على زر الانتقال يظهر الشكل الآتي :



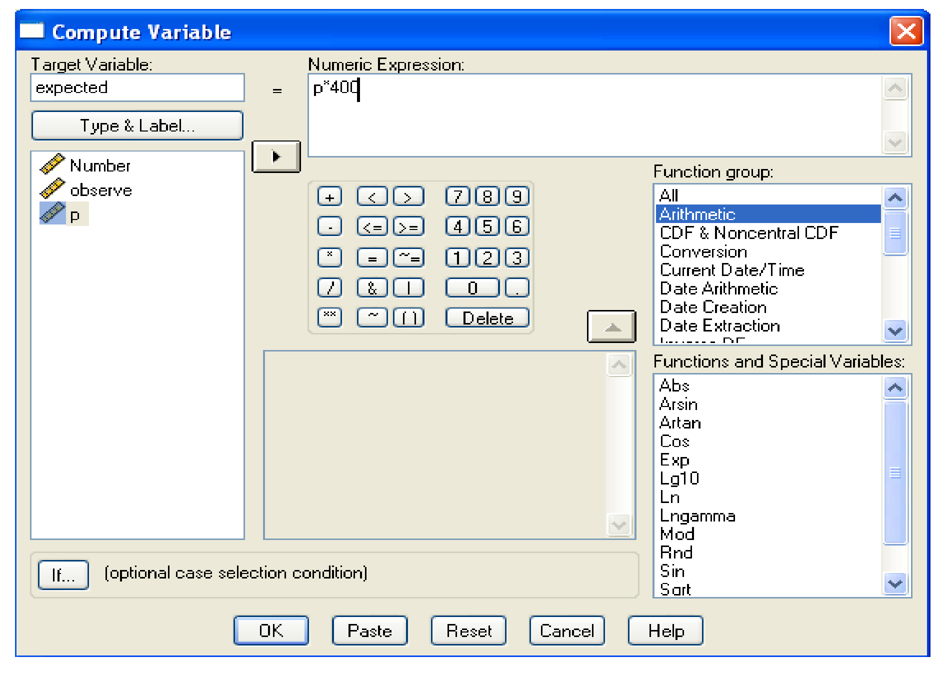
شكل (٤)

ففي علامة الاستفهام الأولى نقوم بإدخال المتغير Number من زر إدخال المتغيرات بعد ذلك نقوم بإدخال عدد مرات المحاولة وهي في المثال تساوي ٣ و أخيرا في علامة الاستفهام الثالث و الأخيرة نقوم بإدخال احتمال الحدوث أو احتمال النجاح وهو في المثال ٠.٥. ونقوم تحت عنوان Target variable نقوم بإدخال اسم المتغير الجديد و هنا هو p الذي يرمز لدالة الاحتمال ومن ثم النقر على ok فتظهر شاشة Data View كالآتي:

جدول (٣)

Number	observe	p
.00	48.0000	.13
1.00	147.000	.38
2.00	153.000	.38
3.00	152.000	.13

ومرة أخرى نقوم بحساب $np_i = 400\,p_i$ بالطريقة السابقة ونسمي المتغير الجديد expected كما في الشكل الآتي:



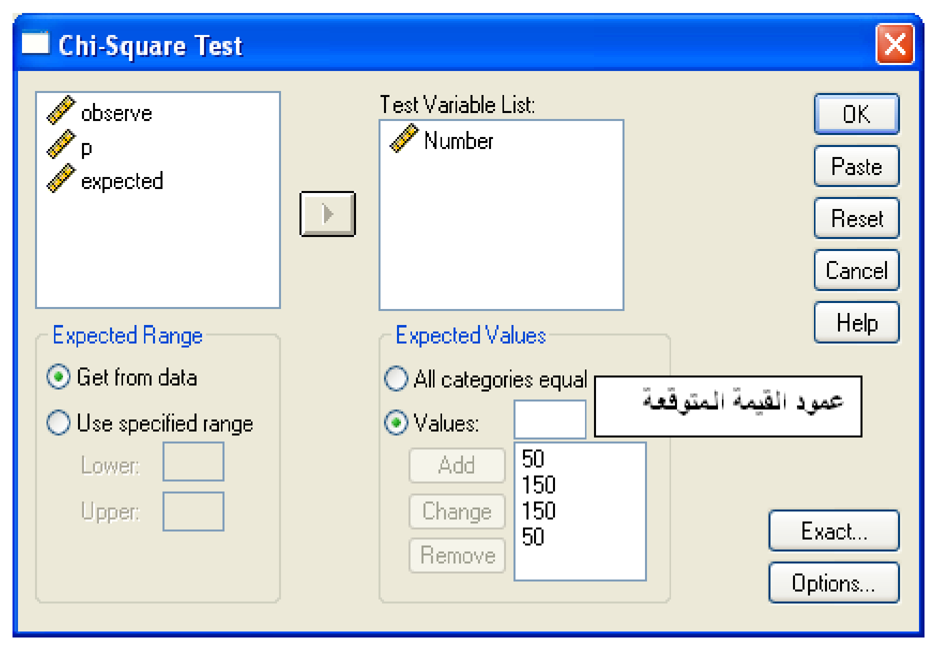
شكل (٥)

بعد ذلك ننتقل إلى الخطوة الأخيرة حيث نقوم بتعريف أن العمود observed هو عمود تكرار العمود Number وذلك من خلال اختيار قائمة Data من شريط القوائم ومن ثم اختيار Weight cases ثم نقوم بقياس العمود الثاني كما في الشكل التالي:

Weight Cases
Number
p
expected
Do not weight cases
Weight cases by
Frequency Variable:
observe
Current Status: Do not weight cases
OK
Paste
Reset
Cancel
Help

شكل (٦)

وبعد ذلك نقوم بفتح Analysis ثم Nonparametric ثم Chi square فيظهر الشكل الآتي:



شكل (٧)

فنقوم بإدخال المتغير Number في عمود list variable test ونقوم بإدخال القيم المتوقعة المحسوبة في العمود الأخير في صفحة Data viewer على التوالي كما في الشكل الآتي حيث نقوم بتنشيط الخيار values من عمود Expected values و بعد ذلك نقوم بإدخال القيمة الأولى ثم ننقر على add ثم الثانية و هكذا و أخيرا نقوم بالنقر على ok في الشكل (٧) وتخرج النتائج التالية:

جدول (٤)

Number

	Observed N	Expected N	Residual
.00	48	50.0	-2.0
1.00	147	150.0	-3.0
2.00	153	150.0	3.0
3.00	52	50.0	2.0
Total	400		

جدول (٤)

Test Statistics

	Number
Chi-Square[a]	.280
df	3
Asymp. Sig.	.964

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 50.0.

ويوضح الجدول (٣) لمربع كاي القيمة المتوقعة لكل خلية.

القرار الإحصائي:

يوضح الجدول (٤) قيمة الاختبار = 0.28 أيضا p-value = .964 وهي أكبر من مستوى المعنوية 0.05 إذا نقبل H_0 بأن عدد الأطفال الذكور بكل أسرة لديها ثلاثة أطفال له توزيع ذي الحدين بنجاح $\theta = 0.5$ وذلك عند مستوى معنوية مقدارها ٠.٠٥.

مثال (٢)

اختيرت عينة عشوائية مكونة من ٥٠٠ طالب من طلبة الفرقة الثالثة بإحدى كليات التجارة وتم تصنيفهم حسب التخصص ونتيجة الامتحان في الإحصاء فكان التصنيف كما يلي في الجدول الآتي:

جدول (5)

	ناجح P	راسب F	Total
إدارة الأعمال	240	60	300
اقتصاد	120	80	200
Total	360	140	500

اختبر ما إذا كان هناك علاقة بين التخصص ونتيجة الامتحان في الإحصاء عند مستوى معنوية 0.05.

الحل:

الفروض الإحصائية:

الفرض العدمي : لا توجد علاقة بين التخصص ونتيجة الإحصاء.

الفرض البديل : توجد علاقة بين التخصص ونتيجة الإحصاء .

⌖ أولا طريقة إدخال البيانات في البرنامج:

نقوم بإدخال رقم الصف الأول و العمود الأول ثم الصف الأول العمود الثاني ثم الصف الثاني العمود الأول ثم الصف الثاني العمود الثاني وبعد ذلك القيم تكون في عمود مستقل، ونعرف العمود الأول x حيث يشير إلى التخصص فيأخذ التخصص إدارة الأعمال (١) والتخصص الاقتصاد (٢) و العمود الثاني y يشير إلى النتيجة في التخصص فتأخذ (١) في حالة النجاح وتأخذ (٢) في حالة الرسوب انظر فصول أساسيات العرض و التحليل الإحصائي باستخدام البرنامج فيظهر الشكل التالي:

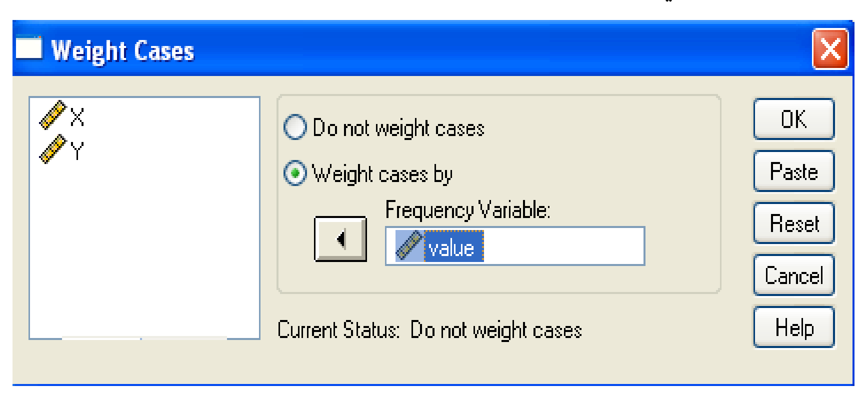
X	Y	value
1	1	240.00
1	2	60.00
2	1	120.00
2	2	80.00

شكل (٨)

ثم نقوم بقياس القيم في العمود الثالث بالطريقة التالية:

Data→ Weight cases

فيظهر الشكل التالي:

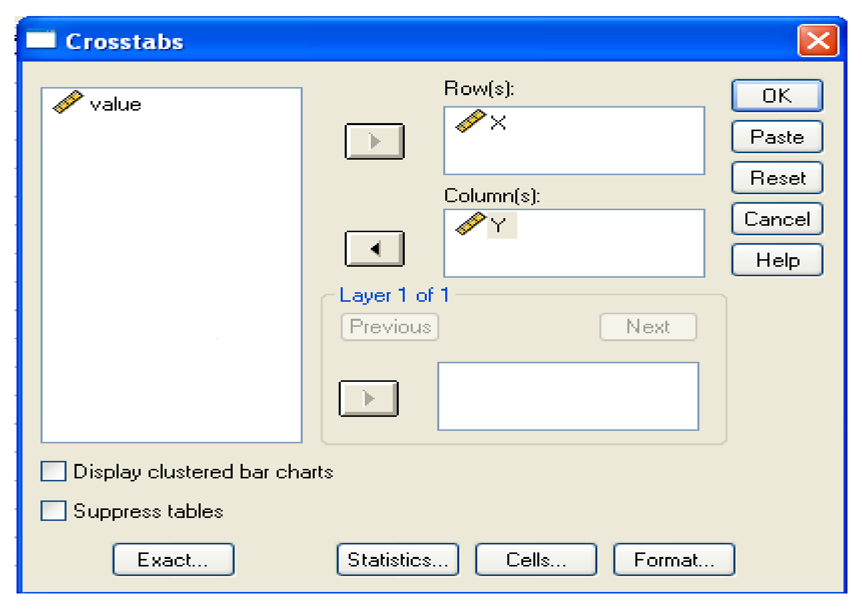


شكل (٩)

ثم نقوم بالتالي:

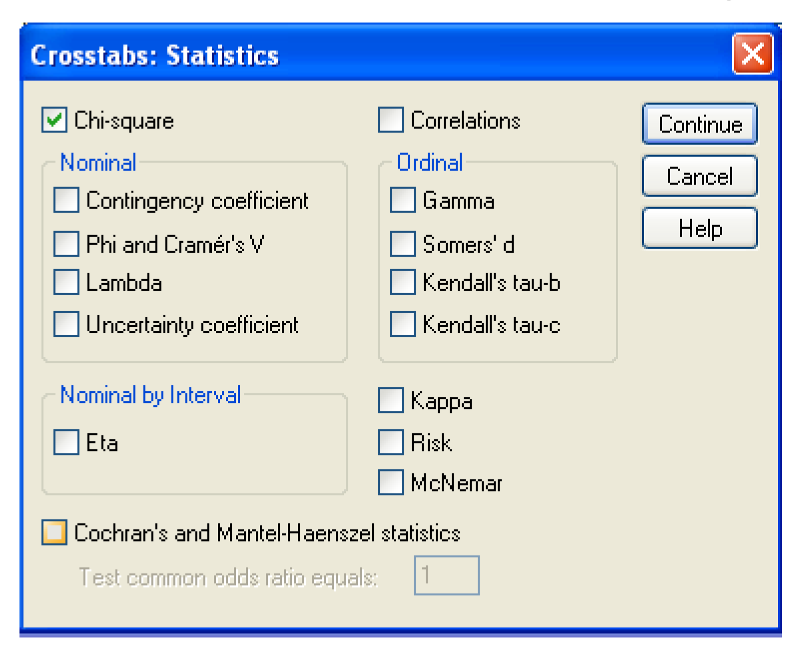
Analyze → Descriptive Statistics → Cross tabs ..

ثم ندخل العمود الأول في Rows و العمود الثاني في column كما يلي:



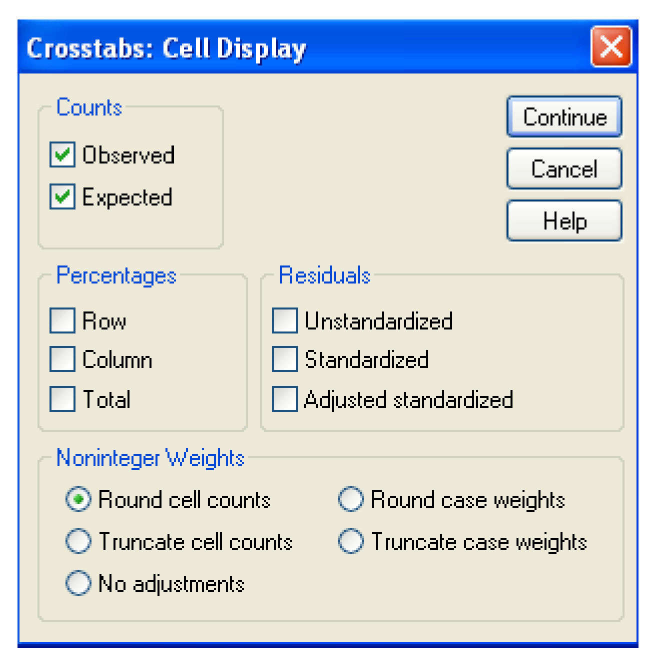
شكل (١٠)

ثم نقوم بالنقر على Statistics ونقوم باختيار Chi-square كما في الشكل التالي ومن ثم النقر على continue لتنشيط :



شكل (١١)

ثم نقوم بالنقر على Cells ثم نقوم باختيار القيم المتوقعة لكل خلية كما في الشكل التالي:



شكل (١٢)

وبعد ذلك ننقر على Continue ثم OK في الشكل (١١) و (12) فنحصل على النتائج التالية:

جدول (٦)

X * Y Crosstabulation

			Y		Total
			1	2	
X	1	Count	240	60	300
		Expected Count	216.0	84.0	300.0
	2	Count	120	80	200
		Expected Count	144.0	56.0	200.0
Total		Count	360	140	500
		Expected Count	360.0	140.0	500.0

جدول (٧)

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	23.810[b]	1	.000		
Continuity Correction[a]	22.828	1	.000		
Likelihood Ratio	23.507	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	23.762	1	.000		
N of Valid Cases	500				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 50.00.

ويوضح جدول (٦) القيمة المتوقعة لكل خلية.

القرار الإحصائي:

يوضح جدول (٧) قيمة الاختبار = 23.810 و أيضا p-value = .000 وهي أقل من مستوى المعنوية 0.05إذا نرفض H_0 ومن ثم توجد علاقة بين التخصص ونتيجة الإحصاء .

مثال (٣)

طرحت مائة فرشة أسنان جديدة على مائة رجل و مائة سيدة لاستخدامها ثم إبداء آرائهم هل يفضلون استخدامها أم لا 32 من الرجال و 26 من السيدات أجابوا بأنهم لا يفضلون استخدام الفرشاة الجديدة. هل هذا يدل على وجود فرق في التفضيل بين الرجال و السيدات عند مستوى معنوية 0.05؟ حيث كانت البيانات كما بالجدول التالي:

جدول (٨)

	يفضل	لا يفضل	Total
رجل	٦٨	٣٢	١٠٠
سيدة	٧٤	٢٦	١٠٠
Total	١٤٢	٥٨	٢٠٠

الحل:

الفروض الإحصائية:

فرض العدم : لا يوجد فرق في التفضيل بين الرجال و السيدات.

الرفض البديل : يوجد فرق في التفضيل بين الرجال و السيدات.

بنفس الطريقة السابقة نحصل على النتائج التالية:

جدول (٩)

VAR00001 * VAR00002 Crosstabulation

			VAR00002		Total
			like	don't like	
VAR00001	man	Count	68	32	100
		Expected Count	71.0	29.0	100.0
	woman	Count	74	26	100
		Expected Count	71.0	29.0	100.0
Total		Count	142	58	200
		Expected Count	142.0	58.0	200.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.874[b]	1	.350		
Continuity Correction[a]	.607	1	.436		
Likelihood Ratio	.875	1	.349		
Fisher's Exact Test				.436	.218
Linear-by-Linear Association	.870	1	.351		
N of Valid Cases	200				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 29.00.

القرار الإحصائي:

في جدول (١١-٩) نلاحظ قيمة اختبار مربع كاي = 0.874 ونلاحظ أيضا قيمة p-value=0.35 وهي أكبر من 0.05 إذا نقبل H_0 ونقول أنه لا يوجد فرق في التفضيل بين الرجال و السيدات في تفضليهم للمنتج.

المراجع

المراجع العربية:

- الهلباوي، عبد الله توفيق، الإحصاء التطبيقي ، مكتبة عين شمس، ٢٠٠٠.
- بشير، سعد زغلول، دليلك إلى البرنامج الإحصائي SPSS،، الإصدار العاشر، الجهاز المركزي للإحصاء جمهورية العراق.
- عاشور،سمير كامل وسالم،سامية أبو الفتوح ، العرض والتحليل الإحصائي بإستخدام SPSSWIN ٢٠٠٣.
- هندي،محمود محمد محمد إبراهيم وسلمان،خلف سلمان سلطان،مفاهيم لطرق التحليل الإحصائي ،مكتبة الرشد،الرياض،١٤٢٥.
- سلطان، عبد الله علي حسن و حسين، علي الهفوفي، اساسيات العرض والتحليل الإحصائي باستخدام SPSS-WIN (الجزء الأول). مشروع تخرج بجامعه الملك سعود، كليه العلوم، قسم الإحصاء وبحوث العمليات، ١٤٢٧ هـ.
- ثامر، محمد عباس منشي ودهام، ناصر دهام الدهام، اساسيات العرض والتحليل الإحصائي باستخدام SPSS-WIN (الجزء الثاني). مشروع تخرج بجامعه الملك سعود، كليه العلوم، قسم الإحصاء وبحوث العمليات، ١٤٢٨ هـ.

المراجع الأجنبية:

- Hollander، M. & Wolf، D. A. (1998). Nonparametric Statistical Methods، Second Edition، Wiley، New York.
- Gupta، V. (1999). SPSS for Beginners. VJBooks Inc.

الفهرس